The Science of Science

Dashun Wang and Albert-László Barabási

Part 1: The Science of Career

This is the submitted version.
The printed version[1] is published by Cambridge University Press in 2021.

[1] Wang, Dashun, and Albert-László Barabási. 2021. *The Science of Science*. Cambridge University Press. doi:10.1017/9781108610834.

Albert Einstein published 248 papers in his lifetime, Charles Darwin 119, Louis Pasteur 172, Michael Faraday 161, Siméon Denis Poisson 158, and Sigmund Freud 330 [1]. Contrast these numbers with the body of work of Peter Higgs, who had published only 25 papers by the age of 84, when he received the Nobel prize for predicting the Higgs boson. Or think of Gregor Mendel, who secured an enduring legacy with only seven scientific publications to his name [2].

These differences show that in the long run what matters to a career is not productivity, but impact. Indeed, there are remarkable differences among the impact of the publications. Even for star scientists, of all papers they publish, at most a few may be remembered by a later generation of scientists. Indeed, we tend to associate Einstein's name with relativity and Marie Curie with radioactivity, while lacking general awareness of the many other discoveries made by each. In other words, one or at most a few discoveries—the outliers—seem to be what define a scientist's career. So, do these outliers accurately represent a scientific career? Or did these superstar scientists just get lucky in one or few occasions along their careers?

And, if only one or at most few papers are remembered, when do scientists make that defining discovery? Einstein once quipped, "A person who has not made his great contribution to science before the age of thirty will never do so." [3]. Indeed, Einstein was merely 26 years old when he published his *Annus mirabilis* papers. Yet, his observation about the link between youth and discovery was not merely autobiographical. Many of the physicists of his generation too made their defining discoveries very early in their career—Heisenberg and Dirac at 24; Pauli, Fermi and Wigner at 25; Rutherford and Bohr at 28. But is youth a necessity for making an outstanding contribution to science? Clearly not. Alexander Fleming was 47 when he discovered penicillin. Luc Montagnier was 51 when he discovered HIV. And John Fenn was 67 when he first began to pursue the research that would later won him the Nobel prize in chemistry. So, how is creativity, as captured by scientific breakthroughs, distributed across the lifespan of a career?

The first part of this book will dive into these sets of fascinating questions regarding scientific careers. Indeed, as we survey our young and not so young colleagues doing groundbreaking work, we are prompted to ask: Are there quantitative patterns underlying when breakthrough work happens in a scientific career? What mechanisms drive the productivity and impact of a scientist? The chapters in this part will provide quantitative answers to these questions, offering insights that affect both the way we train scientists and the way we acknowledge and reward scientific excellence.

Chapter 1.1

Productivity of a scientist

Paul Erdős, arguably the most prolific mathematician in the 20th century, was, by all accounts, rather eccentric. The Hungarian-born mathematician lived out of a ragged suitcase that he famously dragged with him to scientific conferences, universities, and the homes of colleagues all over the world. He would show up unannounced on a colleague's doorstep, proclaim gleefully, "My mind is open." He then spent a few days working with his host, before moving on to surprise some other colleague at some other university. His meandering was so constant that it eventually earned him undue attention from the FBI. To his fellow mathematicians, he was an eccentric but lovable scientist. But to law enforcement officers during the Cold War, it was suspicious that he crossed the Iron Curtain with such ease. Indeed, Erdős was once arrested in 1941 for poking around a secret radio tower. "You see, I was thinking about mathematical theorems," he explained to the authorities in his thick Hungarian accent. It took decades of tracking for the Bureau to finally believe him, concluding that his rambling was indeed just for the sake of math.

His whole *life* was, too. He had no wife, no children, no job, not even a home to tie him down. He earned enough in guest lecturer stipends from universities and from various mathematics awards to fund his travels and basic needs. He meticulously avoided any commitment that might stand in the way of his work. Before he died in 1996 at the age of eighty-three, Erdős had written or co-authored a stunning 1,475 academic papers in collaboration with 511 colleagues. If total publication counts as a measure of productivity, how does Erdős' number compare to the productivity of an ordinary scientist? It surely seems exceptional. But how exceptional?

1.1.1 How much do we publish?

Scholarly publications are the primary mode of communication in science, helping disseminate knowledge. The productivity of a scientist captures the rate at which she adds units of knowledge to the field. Over the past century, the number of publications has grown exponentially. An important question is whether the growth in our body of knowledge is simply because there are now more scientists, or because each scientist produces more on average than their colleagues in the past.

An analysis of over 53 million authors and close to 90 million papers published across all branches of science [4] shows that both the number of papers and scientists grew exponentially over the past century. Yet, while the former grew slightly faster than the latter (Fig. 1.1.1a), meaning that the number of publications per capita has been decreasing over time, for each scientist, individual productivity has stayed quite stable over the past century. For example, the number of papers a scientist produces each year has hovered at around two for the entire 20th century (Fig 1.1.1b, blue curve), and has even increased slightly during the past 15 years. As of 2015, the typical scientist authors or co-authors about 2.5 papers per year. This growth in individual productivity has its origins in collaborations: Individual productivity is boosted as scientists end up on many more papers as co-authors (Fig. 1.1.1b, red curve). In other words, while in terms of how many scientists it takes to produce a paper, that number has been trending downwards over the past century, thanks to collaborative work individual productivity has increased during the past decade.

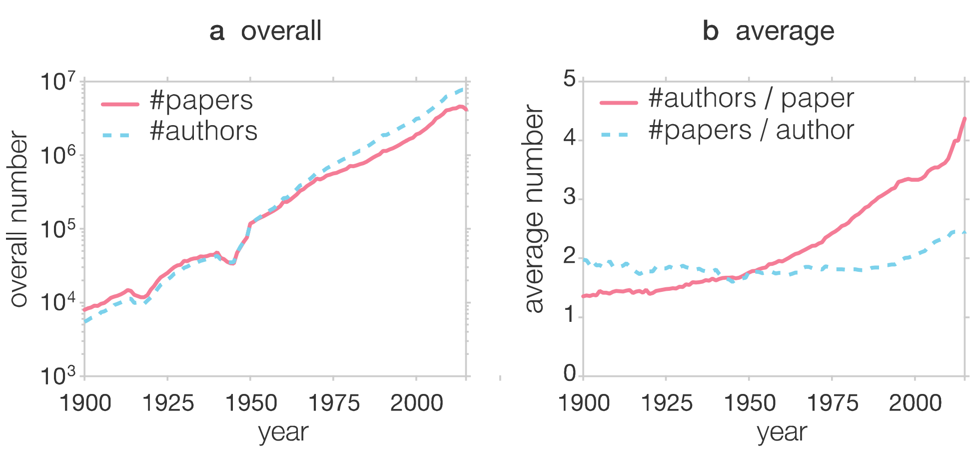

Figure 1.1.1 **The growing number of scientists**. (a) During the past century, both the number of scientists and the number of papers has increased at an exponential rate. (b) The number of papers co-authored by each scientist has been hovering around two during the past 100 years, and increased gradually in the past

15 years. This growth is a direct consequence of collaborative effects: Individual productivity is boosted as scientists end up on many more papers as co-authors. Similar trends were reported using data within a single field [5]. For physics, for example, the number of papers co-authored by each physicist has been less than one during the past 100 years, but increased sharply in the past 15 years. After Dong *et al*. [4] and Sinatra *et al*. [5].

1.1.2 Productivity: Disciplinary ambiguities

But, when it comes to a scientist's productivity, it's not easy to compare across disciplines. First, each publication may represent a unit of knowledge, but that unit comes in different sizes. A sociologist may not feel his theory is fully articulated unless the introduction of the paper spans a dozen pages. Meanwhile, a paper published in *Physical Review Letters*, one of the most respected physics journals, has a strict four-page limit, including figures, tables, and references. Also, when we talk about individual productivity, we tend to count publications in scientific journals. But in some branches of the social sciences and humanities, books are the primary form of scholarship. While each book is counted as one unit of publication, that unit is admittedly much more time-consuming to produce.

And then there is computer science. As one of the youngest scientific disciplines (the first CS department was formed at Purdue University in 1962), computer science has adopted a rather unique publication tradition. Due to the rapidly developing nature of the field, computer scientists choose conference proceedings rather than journals as their primary venue to communicate their advances. This approach has served the discipline well, given everything that has been accomplished in the field—from the Internet to artificial intelligence—but it can be quite confusing to those outside the discipline.

Ignoring the varying publication conventions that characterize different disciplines can have serious consequences. For example, in 2017, the *US News and World Report* (US News), which develops authoritative ranking of colleges, graduate schools, and MBA programs around the world, published their first ranking of the world's best computer science departments. The ranking was so absurd that the Computing Research Association (CRA) had to put out a special announcement, calling it "nonsense" and "a grave disservice" to its readers.

How could an experienced organization specializing in ranking academic institutions get it so wrong? It turns out that *US News* calculated their rankings based on journal publications recorded by Web of Science, a procedure that served them well in all other disciplines. But, by ignoring peer-reviewed papers published in conferences, the *US News* rankings were completely divorced from computer scientists' own perceptions of quality and impact.

The productivity difference across disciplines can be quantified using data from the National Research Council on research doctorate programs in the US [6, 7]. Using the average number of publications by faculty in each department over a five-year period as a proxy, researchers find that the numbers ranged from 1.2 in history to 10.5 in chemistry. Even between similar disciplines we see large productivity differences. For example, within biological sciences, faculty productivity ranged from 5.1 in ecology to 9.5 in pharmacy.

Taken together, the data presented so far in this chapter make at least one message crystal clear: no matter how we measure it, the productivity of a typical scientist is nowhere near Erdős'. Indeed, his total—1,475 papers—implies a staggering *two papers per month over a span of 60 years*. By contrast, a study focusing on more than 15 million scientists between 1996 and 2011, found that less than 1% of our colleagues managed to publish at least one paper every year [8]. Hence, only a small fraction of the scientific workforce can maintain a steady stream of publications. Interestingly, this small fraction contains the most high-impact researchers. Though they represent less than 1% of all publishing scientists, this stable core puts out 41.7% of all papers, and 87.1% of all papers with more than 1,000 citations. And if a productive scientist's pace lags, so does the impact of their contributions. Indeed, the average impact of papers published by a researcher is substantially lower if he skipped even a single year.

While Erdős is an outlier, his impressive productivity speaks to the enormous productivity differences among researchers. Why are there such differences? After all, we all have a 24-hour day to work with. So how can people like Erdős be so much more productive than their peers? To answer these questions, we need to visit the legendary Bell Laboratory in its heyday.

1.1.3 Productivity: The difference

The career of William Shockley, the man who brought silicon to Silicon Valley, was not free of controversies. To be sure, his attempts to commercialize a new transistor design in the 1950s and 1960s transformed the Valley into the hotbed of electronics. Yet, his troubling advocacy for eugenics eventually isolated him from his colleagues, friends, and family. Shockley spent his most productive years at the Bell

Laboratory, where he co-invented the transistor with John Bardeen and Walter Brattain. That discovery not only won the trio the 1956 Nobel prize in Physics, it also began the digital revolution we continue to experience today.

While managing a research group at Bell Labs, Shockley became curious [9]: Were there measurable differences in the productivity of his fellow researchers? So he gathered statistics on the publication records of employees in national labs such as Los Alamos and Brookhaven. Once he charted the numbers, he was surprised by the outcome: The curve indicated that individual productivity, the number of papers published by a researcher, *N*, follows a lognormal distribution

$$P(N) = \frac{1}{N\sigma\sqrt{2\pi}} e^{-\frac{(InN-\mu)^2}{2\sigma^2}}. \tag{1.1.1}$$

Lognormal distributions are fat-tailed, capturing great variations in productivity. In other words, Shockley learned that most researchers publish very few papers, whereas a non-negligible fraction of scientists are orders of magnitude more productive than the average. Evidence for (1.1.1) is shown in Fig. 1.1.2, plotting the distribution of the number of papers written by all authors listed in INSPECT, together with a lognormal fit [10].

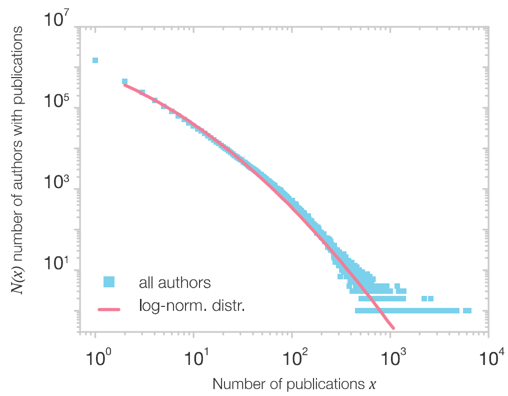

Figure 1.1.2 **Productivity Distribution**. The blue symbols show the number of papers published by all authors listed in the INSPECT database of scientific and technical literature, in the period of 1969-2004 (over 3 million authors). The red line corresponds to the lognormal fit to the data (1.1). After Fronczak *et al.* [10].

Box 1.1.1 The study of productivity has a long history [9-15]

In 1926, Alfred J. Lotka [11] observed that the number of papers produced by scientists follows a fat-tailed distribution. In other words, he found that a small fraction of scientists are responsible for the bulk of scientific literature. Lotka studied 6,891 authors listed in Chemical Abstracts publishing between 1907—1916, concluding that the number of authors making N contributions follows a power law

$$P(N) \sim N^{-\alpha}, \qquad (1.2)$$

where the exponent $\alpha \approx 2$. A power law predicts that productivity has a long tail, capturing major variations among individuals. Note that it often requires a large amount of data to reliably distinguish a power law from a lognormal distribution [9], which Lotka did not have in 1926.

This lognormal distribution of productivity is rather odd, as Shockley quickly noticed. Indeed, in most competitive arenas, individual performance metrics almost always follow a narrow distribution. Think about running. At the Rio Olympics in 2016, Usain Bolt finished the 100-meter final in just 9.81 seconds. Justin Gatlin came in second and Andre De Grasse in third, with running times 9.89s and 9.91s, respectively. These numbers are awfully close, reflecting a well-known fact that performance differences between individuals are typically bounded [16]. Similarly, Tiger Woods, even on his best day, only took down his closest contenders by few strokes, and the fastest typist may only type a few words more per minute than a merely good one. The bounded nature of performance reminds us that it is difficult, if not impossible, to significantly outperform the competition in any domain. Yet, according to Fig. 1.1.2, this boundedness does not hold for scientific performance. Apparently, it *is* possible to be much better than your competitors when it comes to churning out papers. Why is that?

1.1.4 Why so productive?

Shockley proposed a simple model to explain the lognormal productivity distribution (Eq. 1.1.1) he observed [9]. He suggested that in order to publish a paper, a scientist must juggle multiple factors, like:

F_1. Identify a good problem

F_2. Make progress with it

F_3. Recognize a worthwhile result

F_4. Make a decision as to when to stop the research and start writing up the results

F_5. Write adequately

F_6. Profit constructively from criticism

F_7. Show determination to submit the paper for publication

F_8. Make changes if required by the journal or the referees

If any of these steps fail, there will be no publication. Let us assume that the odds of a person clearing hurdle F_i from the list above is p_i. Then, the publication rate of a scientist is proportional to the odds of clearing each of the subsequent hurdles, that is $N \sim p_1 p_2 p_3 p_4 p_5 p_6 p_7 p_8$. If each of these odds are independent random variables, then the multiplicative nature of the process predicts that $P(N)$ follows a lognormal distribution of the form (1.1.1).

To understand where the outliers come from, imagine, that Scientist A has the same capabilities as Scientist B in all factors, except that A is twice as good at solving a problem (F_2), knowing when to stop (F_4), and determination (F_7). As a result, A's productivity will be 8 times higher than B's. In other words, for each paper published by Scientist B, Scientist A will publish eight. Hence small differences in scientists' ability to clear individual hurdles can together lead to large variations in overall productivity.

Shockley's model not only explains why productivity follows lognormal distribution, but it also offers a framework to improve our own productivity. Indeed, the model reminds us that publishing a paper does not hinge on a single factor, like having a great idea. Rather, it requires scientists to excel at multiple factors. When we see someone who is hyper-productive, we tend to attribute it to a single exceptional factor. Professor X is really good at coming up with new problems (F1), or conveying her ideas in writing (F5). The model suggests, however, that the outliers cannot be explained by a single factor; rather, a researcher is most productive when she excels across many factors and fails in none.

The hurdle-model indicates that a single weak point can choke an individual's productivity, even if he or she has many strengths. It also tells us that Erdős may have not been as superhuman as we often think he was, or that his productivity might be attainable with careful honing of various skills. Indeed, if we could improve at every step of writing a paper, and even if it's just a tiny bit in each step, these improvements can combine to exponentially enhance productivity. Admittedly, this is easier said than done. But you can use this list to diagnose yourself: What step handicaps your productivity the most?

The remarkable variations in productivity has implications for reward. Indeed, Shockley made another key observation: while the productivity of a scientist is multiplicative, his salary—a form of reward often tied to performance—is additive. The highest-paid employees earn at best about 50%--100% more than their peers. There are many reasons why this is the case—it certainly seems fairer, and it helps ensure a collaborative environment. Yet, from a paper-per-dollar perspective, Shockley's findings raise some interesting questions about whether the discrepancy between additive salaries and multiplicative productivities could be exploited. Indeed, an institution may be better off employing a few star scientists, even if that means paying them a great deal more than their peers. Shockley's arguments are often used to rationale why top individuals at research-intensive institutions are offered much higher salaries and special perks, and why top departments within a university get disproportionately more funding and resources.

To be sure, gauging a career based on publication count alone grossly misrepresents how science works. Yet, individual productivity has been shown to closely correlate with the eminence of a scientist as well as her perceived contributions to the field. This pattern was documented by Wayne Dennis, dating back at least to 1954 [1], when he studied 71 members of the US National Academy of Sciences and eminent European scientists. He found that, almost without exception, highly productive individuals have also achieved scientific eminence, as demonstrated by their listing in the *Encyclopedia Britannica* or in histories of important developments they have contributed to the sciences. Higher productivity has been shown to increase the odds of receiving tenure [17], and of securing funding for future research [18]. At the institutional level, the publication rates of the faculty are not only a reliable predictor of a program's reputation, they also influence the placement of graduates into faculty jobs [19].

In sum, sustained high productivity is rare, but it correlates with scientific impact and eminence. Given this evidence, it may appear that productivity is the key indicator for a meaningful career in science. Yet, as we show in the following chapters, among the many metrics used to quantify scientific excellence, productivity is the least predictive. The reason is simple: While great scientists tend to be very productive, not all scientists who are productive make long-lasting contributions. In fact, most of them do not. Multiple paths can lead to achieving high productivity. For example, lab technicians in certain fields may find their names on more than a hundred—or sometimes as many as a thousand—papers. Hence, they appear to be exceptionally prolific based on their publication counts, but are rarely credited as the intellectual owner of the research. The way people publish is also changing [20]. Co-authorship is on the rise, as are multiple publications on the same data. There have also been more discussions about LPUs, which stands for Least Publishable Unit [20] or the "salami publishing" approach, which could further contribute to an inflated productivity counts.

So, if productivity is not the defining factor of a successful career, what is?

Box 1.1.2 Name Disambiguation

Our ability to accurately track individual productivity relies on our skill to identify the individual(s) who wrote a paper and all other work that belongs to that individual [21, 22]. This seemingly simple task represents a major unsolved problem [21-23], limited by four challenges. First, a single individual may appear in print under multiple names because of orthographic and spelling variants, misspellings, name changes due to marriage, religious conversion, gender reassignment, or the use of pen names. Second, some common names can be shared by thousands of individuals. Third, the necessary metadata is often incomplete or missing. This includes cases where publishers and bibliographic databases failed to record authors' first names, their geographical locations, or other identifying information. Fourth, an increasing percentage of papers is not only multi-authored, but also represent multi-disciplinary and multi-institutional efforts. In such cases, disambiguating some of the authors does not necessarily help assign the remaining authors.

While multiple efforts are underway to solve the name disambiguation problem, we need to be somewhat mindful about the results presented in this and following chapters, as some conclusions may be affected by the limitations in disambiguation. In general, it is easier to disambiguate productive scientists, who have a long track-record of papers, compared with those who have authored only a few publications. Therefore, many studies focus on highly productive scientists with unusually long careers instead of "normal" scientists.

Chapter 1.2

The H Index

Lev Landau, a giant of Russian physics, kept a handwritten list in his notebook, ranking physicists on a logarithmic scale of achievement and grading them into "leagues" [24]. According to Landau, Issac Newton and Albert Einstein belonged to the highest rank, above anyone else: he gave Newton the rank 0 and Einstein a 0.5. The first league, a rank of 1, contains the founding fathers of quantum mechanics, scientists like Niels Bohr, Werner Heisenberg, Paul Dirac and Erwin Schrödinger. Landau originally gave himself a modest 2.5, which he eventually elevated to 2 after discovering superfluidity, an achievement for which he was awarded the Nobel Prize. Landau's classification system wasn't limited to famous scientists, but included everyday physicists, who are given a rank of 5. In his 1988 talk "My Life with Landau: Homage of a 4 1/2 to a 2," David Mermin, who co-authored the legendary textbook "Solid State Physics," rated himself a "struggling 4.5" [25].

When scientists leave league 5 behind and start approaching the likes of Landau and other founders of a discipline, it's obvious that their research has impact and relevance. Yet for the rest of us, things are somewhat blurry. How do we quantify the cumulative impact of an individual's research output? The challenge we face in answering this question is rooted in the fact that an individual's scientific performance is not just about how many papers one publishes, but a convolution of productivity and impact, requiring us to balance the two aspects in a judicious manner.

Of the many metrics developed to evaluate and compare scientists, one stands out in its frequency of use: the *h* index, proposed by Jorge E. Hirsch in 2005 [26]. What is the *h* index, and how to calculate it?

Why is it so effective in gauging scientific careers? Does it predict the future productivity and impact of a scientist? What are its limitations? And how do we overcome these limitations? Understanding these questions is the aim of this chapter.

1.2.1 The *h*-index

The index of a scientist is *h* if *h* of her papers have at least *h* citations and each of the remaining papers have less than *h* citations [26]. For example, if a scientist has an *h* index of 20 ($h = 20$), it means that she has 20 papers with more than 20 citations, and the rest of her papers all have less than 20 citations. To measure *h*, we sort an individual's publications based on her citations, going from the most cited paper to the least cited ones. We can plot them on a figure, that shows the number of citations of each paper, resulting in a monotonically decreasing curve. Figure 1.2.1 uses the careers of Albert Einstein and Peter Higgs as case studies showing how to calculate their *h*-index.

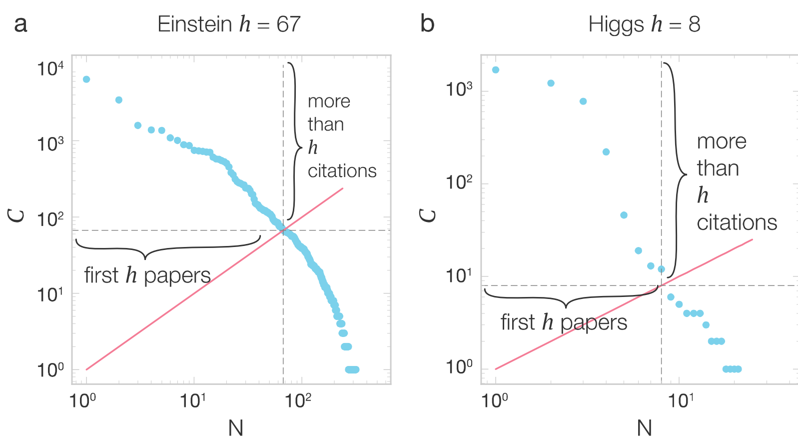

Figure 1.2.1 **The h-index of Albert Einstein (a) and Peter Higgs (b).** To calculate the *h* index, we plot the number of citations versus paper number, with papers listed in order of decreasing citations. The intersection of the 45° line with the curve gives *h*. The total number of citations is the area under the curve [26]. According to Microsoft Academic Graph, Einstein has an *h* index of 67, and Higgs 8. Top three most cited papers by Einstein are 1) Can quantum mechanical description of physical reality be considered complete, Physical Review, 1935; 2) Investigations on the Theory of Brownian Movement, Annalen der Physik, 1905; and 3) On the Electrodynamics of Moving Bodies, Annalen der Physik, 1905. Top three for Higgs are: 1) Broken Symmetries and the Masses of Gauge Bosons, Physical Review Letters, 1964; 2) Broken symmetries, massless particles and gauge fields, Physics Letters, 1964; 3) Spontaneous Symmetry Breakdown without Massless Bosons, Physical Review, 1966.

Is an h index of 8, for example, impressive or modest? What is the expected h index of a scientist? To answer these questions, let's take a look at a simple but insightful model proposed by Hirsch [26]. Imagine that a researcher publishes n papers each year. Let us also assume that each paper earns c new citations every year. Hence a paper's citations increase linearly with its age. This simple model predicts the scientist's time dependent h-index as

$$h = \frac{c}{1+c/n} t. \tag{1.2.1}$$

Therefore, if we define

$$m \equiv \frac{1}{1/c + 1/n}, \tag{1.2.2}$$

we can rewrite (1.1) as

$$h = mt, \tag{1.2.3}$$

indicating that a scientist's h index increases approximately linearly with time. Obviously, researchers don't publish exactly the same number of papers every year (See Chapter 1.1), and citations to a paper follow varied temporal trajectories (as we will cover in Chapter 3.5). Yet, despite the model's simplicity, the linear relationship predicted by (1.2.3) holds up generally well for scientists with long scientific careers [26].

This linear relationship (1.2.3) has two important implications:

(1) If a scientist's h index increases roughly linearly with time, then its speed of growth is an important indicator of her eminence. In other words, the differences between individuals can be characterized by the slope, m. As (1.2.2) shows, m is a function of both n and c. So, if a scientist has higher productivity (a larger n), or if her papers collect more citations (higher c), she has a higher m. And the higher the m, the more eminent is the scientist.

(2) Based on typical values of m, the linear relationship (1.3) also offers a guideline for how a typical career should evolve. For example, Hirsch suggested in 2005 that for a physicist at major research universities, $h \approx 12$ might be a typical value for achieving tenure (i.e., the advancement to associate professor) and that $h \approx 18$ might put a faculty member into consideration for a full professorship. Fellowship in the American Physical Society might typically occur around $h \approx 15$–20, and membership in the US National Academy of Sciences may require $h \approx 45$ or higher.

Since its introduction, the h index has catalyzed a profusion of metrics and greatly popularized the idea of using objective indicators to quantify nebulous notions of scientific quality, impact or prestige [27]. As a testament to its impact, Hirsh's paper, published in 2005, has been cited more than 8,000 times as of the beginning of 2019, according to Google Scholar. It even prompted behavioral changes—some ethically questionable—with scientists adding self-citations for papers on the edge of their *h* index, in hopes of boosting it [28-30]. Given its prevalence, we must ask: can the *h*-index predict the future impact of a career?

Box 1.2.1 The Eddington Number.

The *h* index for scientists is analogous to the Eddington number for cyclists, named after Sir Arthur Eddington (1882-1944), an English astronomer, physicist, and mathematician, famous for his work on the theory of relativity. As a cycling enthusiast, Eddington devised a measure of a cyclist's long-distance riding achievements. The Eddington number, E, is the number of days in your life when you have cycled more than E miles. Hence an Eddington number of 70 would mean that the person in question has cycled at least 70 miles a day on 70 occasions. Achieving a high Eddington number is difficult, since jumping from, say, 70 to 75 may require more than five new long-distance rides. That's because any rides shorter than 75 miles will no longer be included. Those hoping to increase their Eddington number are forced to plan ahead. It might be easy to achieve an E of 15 by doing 15 trips of 15 miles—but turning that E =15 into an E=16 could force a cyclist to start over, since an E-number of 16 only counts trips of 16 miles or more. Arthur Eddington, who reached an E=87 by the time he died in 1944, clearly understood that if he wanted to achieve a high E-number, he had to start banking long rides early on.

1.2.2 The predictive power of the *h* index

To understand the value of the *h* index, let's take a look at the "usual suspects"—metrics that are commonly used to evaluate a scientist's performance, and review their strengths and limitations [26].

i *Total number of publications* (N).

Advantage: Measures the productivity of an individual.

Disadvantage: Ignores the impact of papers.

ii *Total number of citations* (C).

Advantage: Measures a scientist's total impact.

Disadvantage: It can be affected by a small number of big hits, which may not be representative of the individual's overall career, especially when these big hits were coauthored with others. It also gives undue weight to highly cited reviews as opposed to original research contributions.

iii *Citations per paper* (C/N).

Advantage: Allows us to compare scientists of different ages.

Disadvantage: Outcomes can be skewed by highly cited papers.

iv *The number of "significant papers," with more than c citations*.

Advantage: Eliminates the disadvantages of *i*, *ii*, *iii*, and measure broad and sustained impact.

Disadvantage: The definition of "significant" introduces an arbitrary parameter, which favors some scientists or disfavors others.

v *The number of citations acquired by each of the q most-cited papers (for example, q = 5).*

Advantage: Overcomes many of the disadvantages discussed above.

Disadvantage: Does not provide a single number to characterize a given career, making it more difficult to compare scientists to each other. Further, the choice of q is arbitrary, favoring some scientists while handicapping others.

The key advantage of the h index is that it *sidesteps all of the disadvantages* of the metrics listed above. But, is it more effective at gauging the impact of an individual's work? When it comes to evaluating the predictive power of metrics, two questions are often the most relevant.

Q1: Given the value of a metric at a certain time t_1, how well does it predict the value of itself or of another metric at a future time t_2? This question is especially interesting for hiring decisions. For example, if one consideration regarding a faculty hire is the likelihood of the candidate to become a member of the National Academy of Sciences 20 years down the line, then it would be useful to rank the candidates by their projected *cumulative* achievement after 20 years. Hirsch tested *Q1* by selecting a sample of condensed matter physicists and looked at their publication records during the first 12 years of their career and in the subsequent 12 years [31]. More specifically, he calculated four different metrics for each individual based on their career records in the first 12 years, including the h-index (Fig. 1.2.2a), the total number of citations (Fig. 1.2.2b), the total number of publications (Fig. 1.2.2c), and the average number of citations per paper (Fig. 1.2.2d). He then asked if we want to select candidates that have the most total citations by year 24, which one of the four indicators give us the best chance? By measuring the correlation coefficient between future cumulative citations at time t_2 and four different metrics calculated at time t_1, he found that the h index and the number of citations at time t_1 turn out to be the best predictors (Fig. 1.2.2).

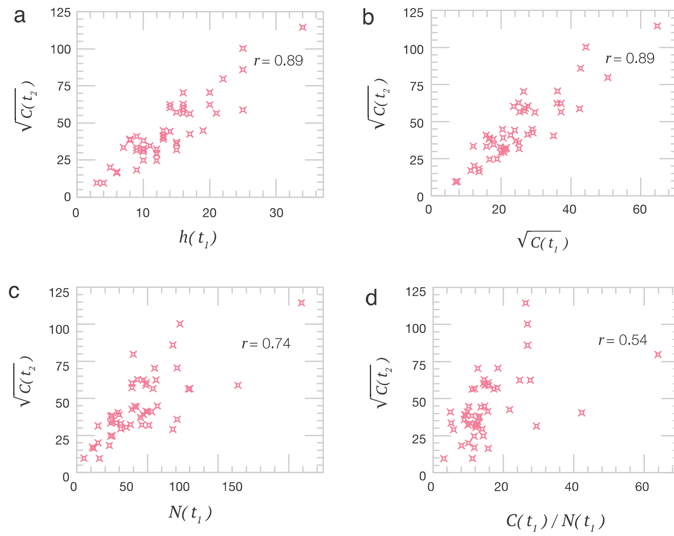

Figure 1.2.2 **Quantifying predictive power of the *h*-index.** Scatter plots compare the total number of citations, c, after $t_2 = 24$ years vs. the value of the various indicators at $t_1 = 12$ year for each individual within the sample. Hirsch hypothesized c may grow quadratically with time, hence used its square root when calculating the total number of citations. By calculating the correlation coefficient, he found that the h index (a) and the number of citations at t_1 (b) are the best predictors of the future cumulative citations at t_2. The number of papers correlates less (c), and the number of citations per paper performs the worst (d). After Hirsch [31].

While Fig. 1.2.2 shows that the h index predicts cumulative impact, in many cases it's the future scientific output that matters the most. For example, if we're deciding who should get a grant, how many more citations an applicant's *earlier* papers are expected to collect in the next few years is largely irrelevant. We're concerned, instead, with papers that the potential grantee has not yet written and the impact of those papers. Which brings us to *Q2*: How well do the different metrics predict *future* scientific output? To answer *Q2*, we need to use indicators obtained at t_1 to predict scientific achievement occurring only in the subsequent period, thereby omitting all citations to work performed prior to t_1. Hirsch repeated the similar prediction task for the four metrics, but this time use each of them to predict total citations accrued by papers published only in the next 12 years. Naturally, this is a more difficult task, but an important one for allocating research resources. Hirsch found that the h index again emerges as the best predictor for achievement incurred purely in future time frame [31].

These findings indicate that two individuals with similar h are comparable in terms of their overall scientific achievement, even if their total number of papers or citations are quite different. Conversely, two individuals of the same scientific age can have a similar number of total papers or citation counts but very different *h* values. In this case, the researcher with the higher *h* is typically viewed by the community as the more accomplished. Together, these results highlight the key strength of the *h* index: When evaluating scientists, it gives an easy but relatively accurate estimate of an individual's overall scientific achievements. Yet at the same time, we must also ask: What are the limitations of the *h* index?

Box 1.2.2 The birth of the *h* index.

Since its inception, the *h* index has been an integral part of a scientific life. Its exceptional influence prompted us to reach out to Jorge Hirsch how he arrived to the measure. He kindly responded, writing:

"I thought about it first in mid-2003, over the next weeks I computed the *h*-index of everybody I knew and found that it usually agreed with the impression I had of the scientist. Shared it with colleagues in my department, several found it interesting.

"mid-June 2005 I wrote up a short draft paper, sent it to 4 colleagues here. one skimmed over it, liked it and made some suggestions, one liked some of it and was nonplussed by some of it, two didn't respond. so I wasn't sure what to do with it.

"mid-July 2005 I got out of the blue an email from Manuel Cardona in Stuttgart saying he had heard about the index from Dick Zallen at Virginia Tech who had heard about it from one of my colleagues at UCSD (didn't say who but I can guess). At that point I decided to clean up the draft and post it in arxiv, which I did August 3rd, 2005, still was not sure what to do with it. Quickly got a lot of positive (and some negative) feedback, sent it to PNAS August 15."

1.2.3 Limitations of the *h*-Index

The main street of College Hill in Easton, Pennsylvania—the home of the Lafayette College—is named after James McKeen Cattell. As an American psychologist, Cattell played an instrumental role in establishing psychology as a legitimate science, advocacy that prompted the *New York Times* to call him "the dean of American science" in his obituary.

While many have thought of developing new metrics to systemically evaluate their fellow researchers, Cattell was the first to popularize the idea of ranking scientists. He wrote in his 1910 book, *American Men of Science: A Biographical Directory* [32]: "*It is surely time for scientific men to apply scientific method to* *determine the circumstances that promote or hinder the advancement of science*." So, today's obsession of measuring impact using increasingly sophisticated yardsticks is by no means a modern phenomenon. Scientists have been sizing up their colleagues since the beginning of the discipline itself. A century after Cattell's book, the need and the rationale for a reliable toolset to evaluate scientists has not changed [33].

As the h index became a frequently used metric of scientific achievements, we must be mindful about its limitations. For example, although a high h is a somewhat reliable indicator of high accomplishment, the converse is not necessarily always true [31]: an author with a relatively low h can achieve an exceptional scientific impact with a few seminal papers, such as the case of Peter Higgs (Fig. 1.2.1b). Conversely, a scientist with a high h achieved mostly through papers with many coauthors would be treated overly kindly by his or her h. Furthermore, there is considerable variation in citation distributions even within a given subfield, and subfields where large collaborations are typical (e.g., high-energy experimental physics) will exhibit larger h values, suggesting that one should think about how to normalize h to more effectively compare and evaluate different scientists.

Next we discuss a few frequently mentioned limitations of the h-index, along with variants that can—at least to a certain degree—remedy them.

- **Highly cited papers.** The main advantage of the h index is that its value is not boosted by a single runaway success. Yet this also means that it neglects the most impactful work of a researcher. Indeed, once a paper's citations get above h, its relative importance becomes invisible to the h index. And herein lies the problem—not only do outlier papers frequently define careers, they arguably are what define science itself. Many remedies have been proposed to correct for this [34-39], including the g-index (the highest number g of papers that *together* received g^2 or more citations [40, 41]) and the o-index (the geometric mean of the number of citations gleaned by a scientist's highest cited papers c^* and her h index: $o = \sqrt{c^* h}$ [42]). Other measures proposed to correct this bias include a-index [36, 38]; $h(2)$-index [39]; hg-index [34]; q^2-index [37]; and more [35].
- **Inter-field differences.** Molecular biologists tend to get cited more often than physicists who, in turn, are cited more often than mathematicians. Hence biologists typically have higher h-index than physicists, and physicists tend to have an h-index that is higher than mathematicians. To compare scientists across different fields, we must account for the field-dependent nature of citations [43]. This can be achieved by the h_g index, which rescales the rank of each paper n by the average number of

papers written by author in the same year and discipline, n_0 [43] or the h_s index, which normalizes h index by the average h of the authors in the same discipline [44].

- **Time dependence.** As we discussed in Ch. 1.2.2., the h-index is time dependent. When comparing scientists in different career stages, one can use the *m*-quotient (1.1.2) [26], or contemporary *h*-index [45].
- **Collaboration effects.** Perhaps the greatest shortcoming of the *h* index is its inability to discriminate between authors that have very different coauthorship patterns [46-48]. Consider two scientists with similar *h*-indices. The first one is usually the intellectual leader of his/her papers, mostly coauthored with junior researchers, whereas the second one is mostly a junior author on papers coauthored with eminent scientists. Or consider the case where one author always publishes alone whereas the other one routinely publishes with a large number of coauthors. As far as the *h* index is concerned, all these scientists are indistinguishable. Several attempts have been proposed to account for the collaboration effect, including fractionally allocating credit in multi-authored papers [48-50], and counting different roles played by each coauthor [51-54] by for example differentiating the first and last authorships. Hirsch himself has also repeatedly acknowledged this issue [46, 47], and proposed the h_α index [47] to quantify an individual's scientific leadership for their collaborative outcomes. Among all the papers that contribute to the *h*-index of a scientist, only those where he or she was the most senior author (the highest *h*-index among all the coauthors) are counted toward the h_α index. This suggests that a high h index in conjunction with a high h_α/h ratio is a hallmark of scientific leadership [47].

In addition to these variations of the *h* index, there are other metrics to quantify the overall achievement of individual scientists, including the *i*10-index, used exclusively by the Google Scholar [55], which computes the number of articles with at least ten citations each; or the SARA method [56], which uses a diffusion algorithm that mimics the spreading of scientific credits on the citation network to quantify an individual's scientific eminence. Despite the multitude of metrics attempting to correct the shortcomings of the *h*-index, to date no other bibliometric index has emerged as preferable to the *h*-index, cementing the status of the *h* index as a widely used indicator of scientific achievement.

As we dug deeper into *h* index and the voluminous body of work motivated by it, it is easy to forget a perhaps more important point: No scientist's career can be summarized by a single number. Any metric, no matter how good it is at achieving its stated goal, has limitations that must be recognized before it is used to draw conclusions about a person's productivity, the quality of her research, or her scientific impact. More importantly, a scientific career is not just about discoveries and citations. Rather, scientists are involved in much broader sets of activities including teaching, mentoring, organizing scientific meetings, reviewing, and serving on editorial boards, to name a few. As we encounter more metrics for scientific eminence, it's important to keep in mind that, while they may help us understand certain aspects of scientific output, none of them alone can capture the diverse contributions scientists make to our community and society [57, 58]. Just as Einstein cautioned: "*Many of the things you can count, don't count. Many of the things you can't count, do count.*"

Therefore, we must keep in mind that the *h* index is merely a proxy to quantify scientific eminence and achievement. But the problem is, in science, status truly matters, influencing the perception of quality and importance of one's work. That's what we will focus on in the next chapter, asking if and when status matters, and by how much.

Chapter 1.3

The Matthew Effect

Lord Rayleigh is a giant of physics, with several laws of nature carrying his name. He is also known beyond the profession thanks to Rayleigh scattering, which answers the proverbial question, "Why is the sky blue?" Rayleigh was already a respected scientist when, in 1886, he submitted a new paper to the *British Association for the Advancement of Science* to discuss some paradoxes of electrodynamics. The paper was promptly rejected on the ground that it did not meet the journal's expectation of relevance and quality. Yet, shortly after the decision, the editors reversed course. Not because anything changed about the paper itself. Rather, it turns out that Rayleigh's name had been inadvertently omitted from the paper when it was first submitted. Once the editors realized it was Rayleigh's work, it was immediately accepted with profuse apologies [59, 60]. In other words, what was initially viewed the scribblings of some "paradoxer," suddenly became worth publishing once it became clear that it was the work of a world-renowned scientist.

This anecdote highlights a signaling mechanism critical in science: the role of scientific reputation. Robert K. Merton in 1968 [60] called this the Matthew effect after a verse in the biblical Gospel of Matthew pertaining to Jesus' parable of the talents: *For to everyone who has will more be given, and he will have an abundance. But from the one who has not, even what he has will be taken away.* The Matthew effect as a concept has been independently discovered in multiple disciplines over the last century, and we will encounter it again in Chapter 3.3, when we discuss citations. In the context of careers, the Matthew effect implies that a scientist's status and reputation alone can bring additional attention and recognition. This means that status not only influences the community's perception of the scientist's credibility, playing an important role in how her work is evaluated, but it also translates into tangible assets—from research funding to access to outstanding students and collaborators—which in turn further improve her reputation. The goal of this chapter is to unpack the role of the Matthew effect in careers. When does it matter? And to what extent?

1.3.1 What's in a name?

The Internet Engineering Task Force (IETF) is a community of engineers and computer scientists who develop the protocols that run the Internet. To ensure quality and functionality, engineers must submit all new protocols as manuscripts that undergo rigorous peer review. For a while, each manuscript included the name of every author. However, beginning in 1999, some manuscripts replaced the full author list with a generic "*et al.*," concealing the name of some authors from the review committee.

By comparing cases where well-known authors were hidden by the *et al.* label with those where the hidden names were little-known, researchers effectively conducted a real-world Lord Rayleigh experiment [61]. They found that when an eminent name was present on a submission, like the chair of a working group, which signals professional standing, the submission was 9.4% more likely to be published. However, the "chair effect" declined by 7.2% when the senior author's name was masked by the *et al.* label. In other words, name-based signaling accounts for roughly 77% of the benefits of having an experienced author as a coauthor on the manuscript.

Interestingly, when the analysis was restricted to a small pool of manuscripts that were "pre-screened," or closely scrutinized, the author name premium disappeared. This means that the status effect only existed when the referees were dealing with high submission rates and low odds of acceptance. In other words, when the reviewers do actually read the manuscript, carefully judging their content, status signals tend to disappear.

Given the exponential growth of science, we frequently encounter the "too many to read" situations. Yet, typically peer review is a rather involved process, with multiple rounds of communication between authors and expert reviewers, suggesting that the status signaling may be less of a concern for scientific manuscripts. Indeed, through those rebuttals and revisions, an objective assessment of the work is expected to prevail. Yet, as we see next, the status effect is rarely eliminated.

Whether an author's status affects the *perceived* quality of his/her papers has been long debated in the scientific community. To truly assess the role of status, we need randomized control experiments, where the same manuscript undergoes two separate reviews, one in which the author identities are revealed and another in which they are hidden. For obvious ethical and logistical reasons, such an experiment is difficult to carry out. Yet, in 2017, a team of researchers at Google were asked to co-chair the program of the 10th Association for Computing Machinery International Conference on Web Search and Data Mining (short for WSDM), a highly selective computer science conference with a 15.6% acceptance rate. The researchers decided to use the assignment as a chance to assess the importance of status for a paper's acceptance [62].

There are multiple ways to conduct peer review. The most common is the "single-blind" review, when the reviewers are fully aware of the identity of the authors and the institution where they work, but, the authors of the paper are not privy to the reviewer's identity. In contrast, in "double-blind" review, neither the authors nor the reviewers know each other's identity. For the 2017 WSDM conference the reviewers on the program committee were randomly split into a single-blind and a double-blind group. Each paper was assigned to four reviewers, two from the single-blind group and two from the double-blind group. In other words, two groups of referees were asked to independently judge the same paper, where one group was aware of who the authors were, while the other was not.

Given the Lord Rayleigh example, the results were not surprising: Well-known author—defined as having at least three papers accepted by previous WSDM conferences and at least 100 computer science papers in total—were 63% more likely to have the paper accepted under single-blind review than in double-blind review. The papers under review in these two processes were exactly the same, therefore, the difference in acceptance rate can only be explained by author identify. Similarly, authors from top universities had a 58% increase in acceptance once their affiliation was known. Further, for authors working at Google, Facebook or Microsoft, considered prestigious institutions in computer science, the acceptance rate more than doubled, increasing by 110%.

These results indicate that in science we *do* rely on reputation—captured by both author identity and institution prestige—when we judge a scientific paper. While we may debate whether the use of such information helps us make better or worse decisions, status signaling plays a key role in getting published in prestigious venues.

Box 1.3.1 Double blind vs single blind

If two papers are identical except that one is written by an unknown scientist while the other by a researcher of considerable reputation, they obviously should have an equal chance of being published. Yet, this may only happen if the reviews are double-blind. Isn't this a strong reason for all science to switch to a double-blind review? New evidence shows that this simple question may not have a simple answer.

An experiment performed by *Nature* [63] shows that making double-blind review optional does not solve the problem. *Nature* started to offer the option of double blind review in 2015. Analyzing all papers received between March 2015 and February 2017 by 25 *Nature*-branded journals, researchers found that corresponding authors from less prestigious institutions were more likely to choose double-blind review, presumably as an attempt to correct for pre-existing biases. However, the success rate at both first decision and peer review was significantly lower for double-blind than single-blind reviewed papers.

Take *Nature,* the flagship journal, as an example. Under a single-blind process, the odds of your paper to be sent out for review is 23%. But if you opt for a double-blind review, those odds drop to 8%. The prospects don't improve once the paper is reviewed: The likelihood that your paper is accepted after review is around 44% for single-blind submissions but only 25% for double-blind papers. In other words, referees are more critical if they are unaware of an author's identity. And these differences matter: If you multiply the probabilities of being sent out for review and acceptance after review, the chances of your paper being accepted in *Nature* through the double-blind review process is a staggering 2%. While a single-blind review is still long shot, it has a success rate of 10.1%. In other words, your odds of acceptance drop roughly five-fold by doing nothing more than checking the double-blind box.

A possible explanation for the observed difference is the quality of papers [63]. Indeed, papers by less-known authors or from less prestigious institutions may not report research of the same quality as those coming from experienced authors from elite institutions. This argument is refuted by a randomized experiment conducted at the *American Economic Review (AER)* [64], an influential journal in economics. From May 1987 to May 1989, *AER* ran a randomized experiment on submitted papers, assigning one half to a double-blind review process and the other half to single-blind one. Because the experiment was randomized, there was no quality difference between the two piles. Yet, the acceptance rate was still much lower for papers in the double-blind group.

These results suggest positive discrimination as a more likely explanation: authors with perceived status are subject of the benefit of the doubt, whereas those without the status get extra scrutiny on issues from research design to methodology. Yet, it is difficult to find objective reasons why double-blind review should not be the norm of scientific publishing. Its widespread use would level the playing field, lowering the hold of researchers with status—whether earned through the impact of previous work, or simply through affiliation with highly prestige institutions.

1.3.2 Boosting Impact

A scientist's reputation facilitates the publication of her paper, but does it also affect the long-term impact of their discoveries? While acceptance only requires a momentary consensus among reviewers, its impact is driven by the broader consensus of the scientific community, who may choose to build further on the work, or may simply ignore it. Do papers published by a well-known scientist also enjoy an impact premium?

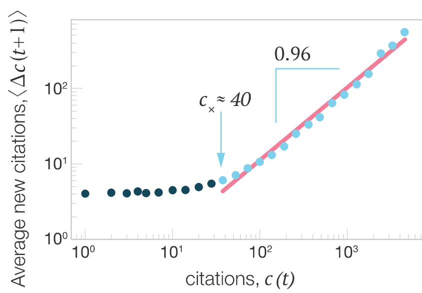

Figure 1.3.1 **The cross-over effect of reputation on citations.** The linear attachment rate breaks down for $c < c_x$, suggesting that additional forces provide a citation boost which elevates $c(t)$ to deviate from what is predicted by the pure preferential attachment mechanism. Datasets include 100 top-cited physicists, and another 100 highly prolific physicists. After Petersen *et al.* [65].

But, how do we measure reputation? A good starting point may be $C_i(t)$, representing the total citation count of all papers an author has published prior to a new paper's publication [65, 66]. Indeed, $C_i(t)$ combines both productivity (the number of papers published by an author) and impact (how often these papers are cited by others), which together offer a reasonable proxy of the author's name recognition within her research community.

The rate at which a paper acquires new citations tends to be proportional to how many citations the paper has already collected [67, 68]. Indeed, highly cited papers are more read, hence are more likely to be cited again. This phenomenon is called preferential attachment, which we will discuss again in Chapter 3.3. It is similar to the Matthew effect that characterizes individual status, but applies to the "status of a paper". To see how an author's reputation affects the impact of her publications, we can look at how their citation patterns deviate from what preferential attachment would predict. This measurement led to the discovery of an early citation premium for well-known authors: Their papers follow preferential attachment only after a certain point. Before that, the number of citations the papers of a prominent author acquire tends to exceed what one would expect, suggesting an early citation premium possibly rooted in the reputation of the author [65]. For example, for a group of well-known physicists, preferential attachment turns on only after the paper has acquired around 40 citations ($c_x \approx 40$) (Fig. 1.3.1). In contrast, for junior faculty in physics (assistant professors), c_x drops from 40 to 10. In other words, right after its publication a senior author's paper is four times more likely to be cited than a junior author's. While its magnitude differs, this effect is also present in other fields: For highly cited cell biologists, $c_x \approx 100$, while for top mathematicians, $c_x \approx 20$.

Figure 1.3.1 suggests that reputation plays an important role early on, when the number of citations is small (i.e., when $c < c_x$). Yet, with time, the reputation effect fads away, and the paper's long-term impact is primarily driven by mechanisms inherent to *papers* rather than their *authors*. In other words, well-known authors enjoy an early citation premium, representing better odds of their work to be noticed by the community. This leads to a leg-up in early citations. But with time, this reputation effect vanishes, and preferential attachment takes over, whose rate is driven primarily by the collective perception of the inherent value of the discovery.

The reputation boost discussed above is not limited to new papers. Eminence can spill over to earlier works as well, boosting their impact. Sudden recognitions, like receiving the Nobel Prize, allow us to quantify this effect. Consider, for example, John Fenn, who received the 2002 Nobel Prize in chemistry for the development of the electrospray ionization technique. His original discovery, published in *Science* in 1989 [69], is Fenn's most cited work, collecting close to 8,500 citations by 2018 according to Google Scholar. But as his landmark paper started to collect citations at an exceptional rate following its publication, the citation rates of several of Fenn's older papers also started to grow at a higher pace. Analyses of 124 Nobel laureates show that this boost is common [70]: The publication of a major discovery increases the citation rates of papers the author published *before*. Interestingly, the older papers that enjoyed the citation boosts are *not necessarily related* to the topic of the new discovery. In other words, reputational signaling operates by bringing professional attention to the individual. Consequently, when an author becomes prominent in one area of science, her reputation is extended to her other line of work, even in unrelated fields.

Box 1.3.2 From Boom to Bust: The Reverse Matthew Effect

If a major breakthrough blesses both past and future scholarship, what does a scandal do to a career? Scientists are certainly fallible, and the scientific community regularly confronts major mistakes or misconduct. These incidents lead to retractions of articles, particularly in top journals [71], where they receive enhanced scrutiny. To what degree does a retracted paper affect a scientific career? Are eminent authors affected more or less severely than their junior colleagues? While retractions are good for science, helping other researchers avoid false hypotheses, retractions are never good for the authors of the retracted paper: they experience a spillover, leading to citation losses to their prior body of work as well [72-74]. The negative impact is not distributed equally, however: Eminent scientists are more harshly penalized for their retracted papers than when retractions happen to their less-distinguished peers [74]. Importantly, this conclusion only holds when the retractions involve fraud or misconduct. In other words, when the retraction is perceived to be the consequence of an "honest mistake," the penalty differential between high- and low-status authors disappears [74].

When a senior and junior scientists are on the same retracted paper, however, the status penalty becomes quite different [75]: Senior authors often escape mostly unscathed, whereas their junior collaborators carry the blame, sometimes even to a career-ending degree. We will return to this effect in Chapter 2.6, where we explore the benefits and the drawbacks of collaborative work.

1.3.3 Is it really the Matthew effect after all?

Great scientists are seldom one-hit wonders [60, 76]. Newton is a prime example: beyond the Newtonian mechanics, he developed the theory of gravitation, calculus, laws of motion, optics, and optimization. In fact, well-known scientists are often involved in multiple discoveries, another phenomenon potentially explained by the Matthew effect. Indeed, an initial success may offer a scientist legitimacy, improve peer perception, provide knowledge of how to score and win, enhance social status, and attract resources and quality collaborators, each of these payoffs further increasing her odds of scoring another win. Yet, there is an appealing alternative explanation: Great scientists have multiple hits and consistently succeed in their scientific endeavors simply because they're exceptionally talented. Therefore, future success again goes to those who have had success earlier, *not* because of advantages offered by the previous success, but because the earlier success was indicative of a hidden talent. The Matthew effect posits that success *alone* increases the future probability of success, raising the question: Does status dictate outcomes, or it simply reflects an underlying talent or quality? In other words, is there really a Matthew effect after all?

Why should we care about which is the more likely explanation, if the outcome is the same? Indeed, independent of the mechanism, people who have previously succeeded are more likely to succeed again in the future. But, if innate differences in talent is the only reason why some people succeed while others don't, it means that the deck is simply stacked in favor of some—at the expense of others—from the outset. If, however, the Matthew effect is real, each success you experience will better your future chances. You may not be Einstein, but if you are lucky to get that early win, you may narrow the gap between yourself and someone of his eminence, as your success snowballs.

Unfortunately, it is rather difficult to distinguish these two competing theories, as they yield similar empirical observations. One test of these contrasting hypotheses was inspired by the French Academy's mythical "41st chair." The Academy decided early on to have only forty seats, limiting its membership to 40 so-called "immortals," and would only consider nominations or applications for new members if one of the seats became vacant through the death of a member. Given this restriction, many deserving individuals were never elected into the Academy, being eternally delegated to the 41st chair. It's a crowded seat, shared by true immortals like Descartes, Pascal, Molière, Rousseau, Saint-Simon, Diderot, Stendahl, Flauberta, Zola, and Proust [60]. At the same time, many of those who did hold a seat in the esteemed club are (unfortunately) utterly irrelevant to us today. With time, the 41st chair became a symbol of the many talented scientists who *should* have been, but were never, recognized as giants of their discipline.

But, does it actually matter if someone is formally recognized or not? Indeed, how does the post-award perception of major prizewinners compare to scientists who had comparable performance, but who were not officially recognized? In other words, how does the career of those that occupied the 41st chair differed, had they been elected to the French Academy? The answer is provided by a study, exploring the impact of a major status-conferring prize [77].

As a prestigious private funding organization for biomedical research in the United States, the Howard Hughes Medical Institute (HHMI) selects "people, not projects," generously supporting scientists rather than awarding them grants for specific peer-reviewed research proposals. The HHMI offers about $1 million per investigator each year, providing long-term, flexible funding that allows awardees the freedom to follow their instincts, and if necessary, change research directions. Beyond the monetary freedom, being appointed an HHMI investigator is seen as a highly prestigious honor. To measure the impact of the HHMI prize, the challenge is to create a control group of scientists who were close contenders but who were not selected for the prize and compare their scientific outputs with those by the HHMI investigators.

But, let's assume that we identify this control group of scientists, and do find evidence that HHMI investigators have more impact. How can we know that the difference is purely because of their newfound status? After all, the $1 million annual grant gives them the resources to do better work. To sort this out, we can focus only on articles written by the prizewinners *before* they received the award. Therefore, any citation differences between the two groups couldn't be simply the result of the superior resources offered to prizewinners. Sure enough, the analysis uncovered a post-appointment citation boost to *earlier* works, offering evidence that in science, the haves are indeed more likely to have more than the have-nots.

This success-breeds-success effect is not limited to HHMI investigators. When a scientist moves from a laureate-to-be to a Nobel Laureate, her previously published work—whether of Nobel prize-winning caliber or not—gathers far more attention [78]. Once again, like the case of John Fenn discussed above, a person's previous work doesn't change when she becomes an HHMI investigator or a Nobel Laureate. But with new accolades casting warm light on her contribution, attention to her work increases.

Interestingly, though, strictly controlled tests suggest that status has only a modest role on impact, and that role is limited to a short window of time. Consistent with theories of the Matthew effect, a prize has a significantly larger effect when there is uncertainty about article quality, and when prizewinners are of (relatively) low status at the time of the award. Together, these results suggest that while the standard approach to estimating the effect of status on performance is likely to overstate its true influence, prestigious scientists do garner greater recognition for their outputs, offering further support for the Matthew effect.

Box 1.3.4 Causal evidence for the Matthew effect: Field experiments.

Randomized experiments offer the best way to untangle the role of status from individual differences such as talent. We can select two groups—a control and a treatment group—and randomly assign an advantage to some while denying it to others. If success is allocated independent of prior success or status, any discrepancy in the subsequent advantage of recipients over non-recipients can only be attributed to the exogenously allocated early success.

While we can't assign life-altering awards or grants to randomly chosen scientists [79], we *can* explore the phenomenon using experiments carried out in real-world settings where the intervention causes minimal harm. This is what Arnout van de Rijt and his collaborators did in a series of experiments [80, 81]. They randomly selected the most productive Wikipedia contributors within a subset of the top 1% of editors and randomly assigned them to one of two groups. Then they gave out "barnstars" to the experimental group—an award used within the community to recognize outstanding editors, while leaving the control group unrecognized. As shown in Fig. 1.3.2, prior to intervention, the activities of the two groups are indistinguishable, as they were drawn randomly from the same sample of productive editors. Yet once the fake barnstars were bestowed on the experimental group, the awardees exhibited more engagement than their peers in the control group, demonstrating greater sustained productivity and less likelihood of discontinuing their editorial activity. Indeed, receiving a barnstar increased median productivity by 60% compared to the control group. Most importantly, they also went on to win many more *real* barnstars from other editors. These additional awards were not just due to their increased productivity, since within the experiment group, the awarded individuals were not more active than those who received no additional barnstars. The observed success-breads-success phenomenon was shown to persist across domains of crowd-funding, status, endorsement, and reputation [81], documenting that initial endowments, even if they are arbitrary, can create lasting disparities in individual success, and offering causal evidence for the Matthew effect.

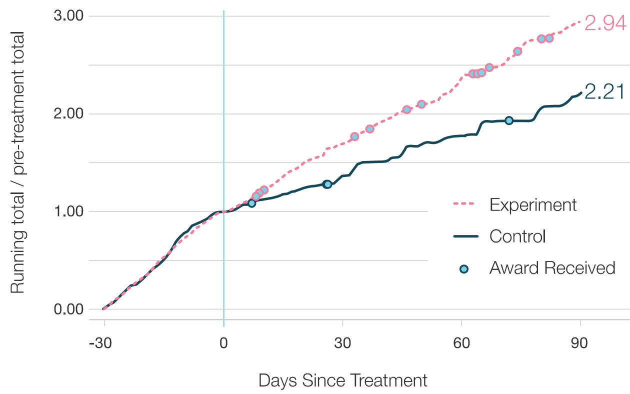

Figure 1.3.2 **The Matthew effect—Evidence from a field experiment.** Researchers randomly assigned Wikipedia editors into two groups, then awarded barnstars to the experiment group, and did nothing for the control group. Circles show when the editors received additional real awards after the treatment. Twelve subjects in the experimental group received a total of fourteen awards, whereas only two subjects in the control condition received a total of three awards. After Restivo *et al.* [80].

Chapter 1.4

Age and Scientific Achievement

When Elias Zerhouni was named director of the National Institutes of Health (NIH) in 2002, he faced a seemingly intractable crisis: The investigators that his organization funds were aging at an alarming rate. Back in 1980, for example, about 18% of all recipients of R01s—NIH's most common research grants for individual investigators—were junior scientists aged 36 or younger. Senior investigators, those 66 years old or above, accounted for less than 1% of the grantees. In the 30 years since, however, a remarkable reversal has taken place (Fig. 1.4.1) [82]: The number of senior investigators increased *ten-fold* by 2010, whereas the percentage of young investigators plummeted dramatically, from 18% to 7%. In other words, back in 1980, for every senior investigator, the NIH funded 18 junior faculty. By 2010, there were twice as many senior investigators as those just starting out. Zerhouni declared this trend to be "the number one issue in American science" [83].

Why was the director of NIH so worried about this trend? After all, senior researchers have a proven track record, know how to manage projects, understand risks, and can serve as experienced mentors for the next generation of scientists. So, when after a rigorous peer review process, senior PIs win out over the fresh-faced youngsters, shouldn't we feel reassured that our tax dollars are in safe hands?

To comprehend the threat this demographic shift poses to the productivity and preeminence of American science, we need to explore a question that has fascinated researchers for centuries: At what age does a person make their most significant contribution to science?

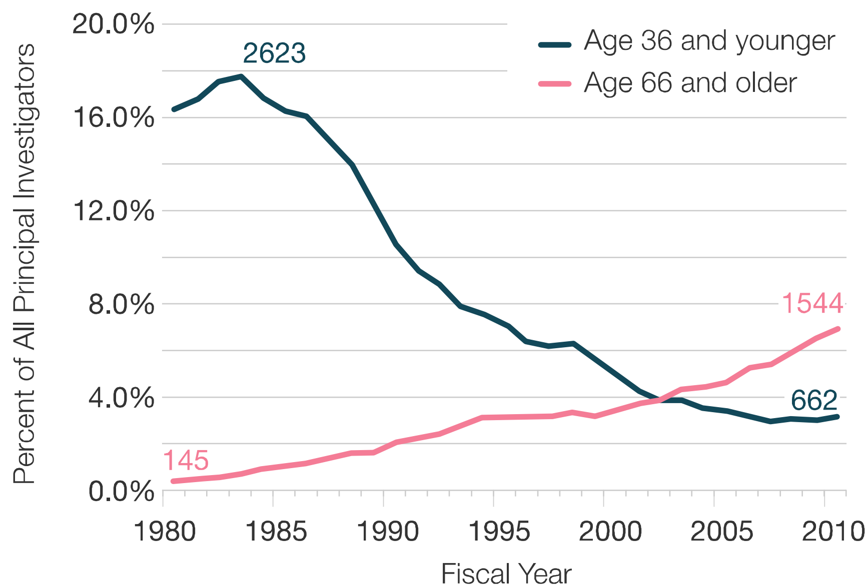

Figure 1.4.1 **The Graying of Science**. Changes in the percentage of NIH R01 grant recipients, aged 36 and younger and aged 66 and older, 1980—2010. After Alberts *et al.* [82].

1.4.1 When do scientists do their greatest work?

The earliest investigation into the link between a person's age and exceptional accomplishment dates back to 1874, when George M. Beard estimated that peak performance in science and the creative arts occurred between the ages of 35 and 40 [84]. Subsequently, Harvey C. Lehman devoted around three decades to the subject, summarizing his findings in *Age and Achievement*, a book published in 1953 [85]. Since then, dozens of studies have explored the role of age in a wide range of creative domains, revealing a remarkably robust pattern: No matter what creative domain we look at or how we define achievement, one's best work tends to occur around mid-career, or between 30 to 40 years of age [2, 66, 85-87].

Figure 1.4.2 shows the age distribution of signature achievements, capturing Nobel Prize winners and great technological innovators of the 20th century [88]. The figure conveys three key messages:

- There is a large variance when it comes to age. While there are many great innovations by individuals in their 30s (42%), a high fraction contributed in their 40s (30%), and some 14% had their breakthrough beyond the age of 50.

- There are no great achievers younger than 19. While Einstein had his *annus mirabilis* at the tender age of 26, and Newton's *annus mirabilis* came even earlier, at the age of 23, the Einsteins and Newtons of the world are actually rare, because only 7% of the sample have accomplished their great achievement at or before the age of 26.
- The Nobel laureates and the inventors come from two independent data sources, with only 7% overlap between the two lists. Yet, the age distributions of these two samples are remarkably similar.

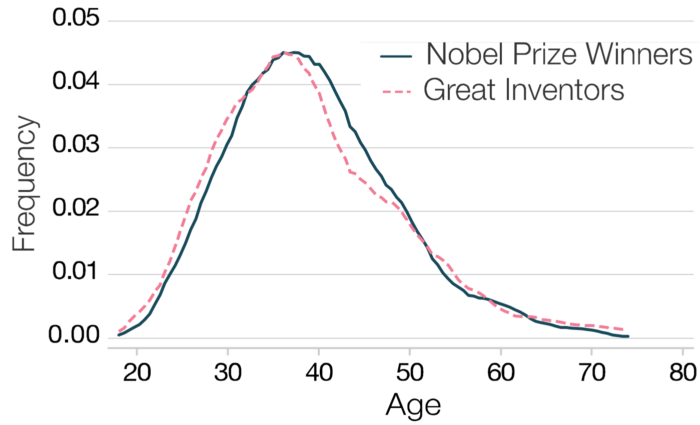

Figure 1.4.2 **Age distribution of great innovation.** The plot shows the age of innovators at the moment of their great innovation, combining all twentieth-century observations. After Jones [88].

Thus, Fig. 1.4.2 demonstrates that scientific performance peaks in middle age [2, 66, 85-87]. The life-cycle of a scientist often begins with a learning period, absent of major creative outputs. This is followed by a rapid rise in creative output that peaks in the late 30s or 40s and ends with a slow decline as he advances through his later years. These patterns are remarkably universal. Researchers have explored them in a variety of ways, identifying important scientists by their Nobel Prizes, by their listings in encyclopedias, and by their membership in elite groups like the Royal Society or the National Academies. No matter how you slice the data, the patterns observed in Fig. 1.4.2 remain essentially the same, raising two questions: Why does creativity take off during our 20s and early 30s? And why does it decline in later life?

1.4.2 The life-cycle of a scientist

The Early Life Cycle. A remarkable feature of a scientific career is the lack of contributions in the beginning of life [89]. Simply put, no 18-year-old has managed to produce anything worthy of a Nobel. The early life-cycle coincides with schooling, suggesting that the need for education may be responsible for the absence of exceptional output in early age. This is also consistent with theories of creativity, where "creativity" is often defined as an ability to identify interesting combinations of existing knowledge [90-92]. Indeed, if we see knowledge as Lego pieces, and new inventions and ideas hinge on figuring out novel ways to combine these pieces, then we first need to accumulate enough pieces before we can start building anything meaningful.

Empirical evidence using World Wars I and II as natural experiments [88], supports that training is responsible for the lack of discovery and invention in the early life cycle. Indeed, war created interruptions that affected the training phase of careers. When Nobel laureates experienced either World War I or II between the ages of 20 and 25, their probability of creating innovative work between the ages of 25 and 30 was substantially reduced, even though the wars were over. This means, scientists do not magically start innovating at a certain age; rather, interruptions during their training phase must be made up. This finding is consistent with the much-discussed "10,000 hours hypothesis" in psychology, which estimates that roughly 10 years of deliberate practice are needed to achieve excellence across broad domains of human performance [93-95].

The Middle and Late Life-Cycle. Once a scientist completes the training phase, a productive career may blossom, as evidenced by the prominent peak in Fig. 1.4.2. Yet, Fig. 1.4.2 also shows that the likelihood of a scientific breakthrough decreases following this peak around middle age. Compared to the rapid rise in the early life cycle, this decline is relatively slower.

There are many potential explanations for this decline, from the obsolescence of skills to degradation in health. But none of them offer a satisfactory explanation for why the decline occurs so early, usually before a scientist reaches 50—long before he is considered old, biologically. For this reason, alternative explanations that consider the realities of the research process stand out as more plausible, including family responsibilities and increasing administrative duties. In other words, older scientists may well have the capacity to produce great science but have limited time to do so. Instead, they're running labs, applying for grants, reviewing papers and deciding tenure cases. Interestingly, scientific careers are sometimes characterized by a “second peak” prior to retirement, which can be interpreted as a rush to get remaining ideas and unpublished research in press [89, 96-98].

These findings help us return to the challenge facing American science (Fig. 1.4.1): While senior researchers have more experience, scientists tend to produce their best work at a younger age, suggesting that a preference for funding only older researchers could stifle innovation. For example, from 1980 to 2010, 96 scientists won a Nobel Prize in medicine or chemistry for work supported by NIH. Yet their prize-winning research was conducted at an average age of 41 [99]—a full year younger than the average age of a NIH investigator starting out today.

1.4.3 What determines the timing of our peak?

What determines when our creativity peaks? Many believe that a scientist’s field is a key factor, as some fields are “young” and others are “old”. Indeed, the prevailing hypothesis is that individuals working in disciplines that focus on deduction and intuition tend to have their great achievements at younger ages, explaining why peak performance comes earlier in disciplines like math or physics than in medicine [85, 100-102].

But this stereotype—of precocious physicists and wizened medical doctors—is increasingly contested. For example, researchers tabulated peak age in different fields by surveying the literature. Upon putting all these numbers together in the same list, they found that the peak age was all over the map [89], lacking obvious so-called young or old fields. Even the canonical example of physics vs medicine doesn’t seem to hold. For example, while it’s true that physics Nobel laureates during the 1920s and 1930s were younger than scientists in all other fields, since 1985, physics laureates have actually been *older* when they made their prizewinning contribution compared to other fields [103].

If field is not a determining factor, then what is? As we discuss next, two convincing theories have emerged: the burden of knowledge and the nature of work.

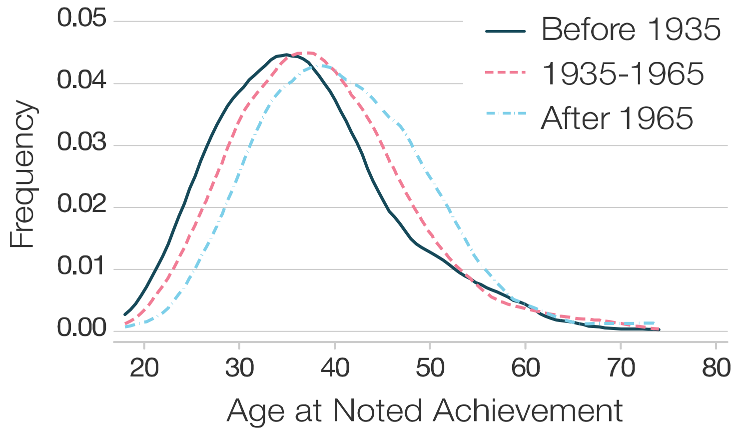

Figure 1.4.3 **Shifts in the Age Distribution of Great Innovation.** Data includes Nobel Prize winners and great inventors. The twentieth century is divided into three chronological periods: 1900–1935, 1935–1965, and 1965 to present. After Jones [88].

Burden of Knowledge

A simple reexamination of Fig. 1.4.2 reveals an interesting pattern [88]. If we take the datasets of Nobel prizewinners and great inventors, and divide the twentieth century into three consecutive chronological periods, we observe a systematic shift (Fig. 1.4.3): The peak age of great minds has increased over time. Indeed, in the early days of the Nobel prize, two thirds of the recipients did their prize-winning work by age 40, and 20 percent earned the distinction before they turned 30. Lawrence Bragg, the youngest Nobel Laureate in Physics, received his award at the astonishing age of 25. Today, a 25-year-old physicist has probably only recently decided on which PhD program to enroll at to begin her training. And since 1980, the mean age for Nobel winning achievement in physics has shifted to 48. Overall, as the 20th century progressed, the great achievements by both Nobel laureates and inventors have occurred at later and later ages, with the mean peak age rising by about six years in total.

There are two plausible explanations for this shift. The first hypothesis is that the life cycle of innovation has changed, so that great minds now innovate at a later stage of their career. This could be due to the extended length of time required for education, which delays the onset of active innovative careers. The second hypothesis reasons that the upward age trend simply reflects general demographic shifts. In other words, if everyone in the world is getting older on average, science as a profession should be no exception.

Yet, even after controlling for demographic effects, substantial shifts in peak age remain unexplained, especially the delays in the early life cycle, when innovators began their active careers [88]. For example, while at the start of the twentieth century a scientist became "research active" at age 23, by the end of the century this moment had shifted to age 31.

To understand the origin of the observed shift, Jones proposed the "burden of knowledge" theory [88, 104, 105]. First, innovators must undertake sufficient education to reach the frontier of knowledge. Second, because science has been growing exponentially, the amount of knowledge one has to master to reach that frontier increases with time. This theory offers a fresh perspective of Newton's famous remark about "standing on the shoulders of giants": In order to stand on a giant's shoulders, one must first climb up his back. But the greater the body of knowledge, the longer the climb.

Clearly, education is an important prerequisite for innovation [88]. Yet if scientists tend to innovate when they are young, then each additional minute they spend in training is one less minute they can spend pushing science forward, potentially reducing the total career output of individual scientists. And if individuals have less time to innovate, the society as a whole innovates less [89]. A back-of-the-envelope calculation suggests that a typical R&D worker contributes approximately only 30% of his time to aggregate productivity gains as he did at the beginning of the 20th century—a shift that can be attributed to the burden of knowledge. In other words, an increase in the age of peak performance potentially contributes to the well documented decline in per capita output of R&D workers in terms of both patent counts and productivity growth [106].

Box 1.4.1 Nobel Prize Threatened.

It's not only discovery that's delayed these days—so is recognition [107]. Indeed, there is a steadily increasing gap in time between when a scientist makes a discovery worthy of a Nobel and when she is awarded for it. Before 1940, only 11% of physics, 15% of chemistry and 24% of physiology or medicine Nobel Prizes recipients had to wait more than 20 years for their recognition after the original discovery. Since 1985, a two-decade delay has affected 60%, 52% and 45% of the awards, respectively. The increasing interval between discovery and its formal recognition is approximated by an exponential curve, projecting that by the end of this century, the prizewinners' average age for receiving the award is likely to exceed their projected life expectancy. In other words, most candidates will not live long enough to attend their Nobel ceremonies. Given that the Nobel prize cannot be awarded posthumously, this lag threatens to undermine science's most venerable institution.

Experimental vs. Conceptual Innovators

By the time Heisenberg entered the University of Munich in 1920, it was clear that the leading theory of the atom developed by Bohr, Sommerfeld, and others, while successful in certain domains, had encountered fundamental challenges. Heisenberg earned his PhD in Munich with Arnord Sommerfeld in 1923 and moved on to work with Max Born in Göttingen, where he developed matrix mechanics in 1925. The following year he became Niels Bohr's assistant in Copenhagen, where, at the age of 25, he proposed the uncertainty principle. In the past century, historians have repeatedly documented his "extraordinary abilities: complete command of the mathematical apparatus and daring physical insight,"—to borrow from Sommerfeld, his PhD advisor.

What is less known, however, is that he nearly failed his PhD defense. On July 23, 1923, the 21-year-old Heisenberg appeared before four professors in the University of Munich. He easily handled Sommerfeld's questions and those relating to mathematics, but he stumbled with astronomy, and flunked badly in experimental physics [108]. When all was said and done, he passed with a C-equivalent grade. When he showed up the next morning at Max Born's office—who had already hired him as his teaching assistant for the coming year—Heisenberg said sheepishly, "I wonder if you still want to have me."

Why, you might ask, did an astonishingly brilliant physicist fail an exam covering the basics of his field? Perhaps because of the kind of innovator he was. There are, broadly speaking, two polar extremes in scholarly creativity: conceptual and experimental [109]. *Experimental innovators* work inductively, accumulating knowledge from experience. This type of work relies heavily on the works of others and tends to be more empirical. Heisenberg was, on the other hand, a *conceptual innovator*, who works deductively, applying abstract principles. Consequently, his work tends to be more theoretical and derives from a priori logic. This distinction between experimental and conceptual creativity suggests that the nature of the work a scientist does determines when she peaks: Conceptual innovators tend to do their most important work earlier in their careers than their experimental colleagues.

Figure 1.4.4 **Theorists vs empiricists among Nobel Laureates.** This figure separates profiles for Nobel laureates, whose prize-winning work was theoretical compared to those whose prize-winning work was empirical, showing clear differences in lifecycle of creativity between the two. After Jones [89].

To explore the distinct life-cycles of the two types of scientists, researchers rated Nobel-winning economists as either conceptual or experimental based on the nature of their discovery, and examined the age when each group published their best work [109]. The difference was stark: Conceptual laureates made their most important contributions to science at the average age of 35.8. Compare that to the average age of 56 for experimental laureates—a staggering difference of 20.2 years. Indeed, 75% of conceptual Nobel laureates in economics published their best work within the first 10 years of their career, while *none* of the experimental laureates managed to do so.

Another way of looking at the nature of work is to separate it into the categories of empirical or theoretical. Note that this distinction is not equivalent with conceptual vs experimental. Indeed, while conceptual innovations tend to be more theoretical, a conceptual innovator can certainly be rooted in empirical work as well. Likewise, experimental innovators can also make theoretical contributions. Nevertheless, when Nobel laureates were separated into either empirical or theoretical categories, a similar pattern emerged (Fig. 1.4.4) [89]: the empirical researchers did their prize-winning work 4.6 years later (at 39.9 years of age compared to 35.3 years) than those doing theoretical work.

There are many reasons why conceptual work tends to happen early and experimental work tends to happen later in a career [89]. First, conceptual innovators like Heisenberg do not need to accumulate large amounts of experience in order to make their contributions. By contrast, Charles Darwin, who did experimental work, needed time to accumulate the evidence necessary to develop his theories. Sure, he was a brilliant naturalist, but his work would have gone nowhere if he hadn't boarded the Beagle to voyage around the world. Collecting all the pieces of evidence and drawing connections among them took time. Indeed, the trip itself took five years. That means it took Darwin half-decade of travel to even have the raw materials for writing *The Voyage of the Beagle*.

Second, some of the most important conceptual work involves radical departures from existing paradigms, and it may be easier to disrupt a paradigm shortly after initial exposure to it, before an individual has produced a large body of work that rests upon the prevailing wisdom. Thus, while experience benefits experimental innovators, newness to a field benefits conceptual innovators. Indeed, nearly failing his PhD exam did not prevent Heisenberg from doing groundbreaking work. To a certain degree, it might have even helped. In other words, knowing too much might even "kill" you as a scientist [110, 111].

Figure 1.4.5 **Michelangelo vs. Picasso.** Two famous paintings side by side. (A) The creation of Adam, by Michelangelo (1511-1512). (B) Woman with Mustard Pot, by Picasso (1910). Can you guess which of these two artists' career peaked early and which is the one that required longer time to mature?

In some sense, conceptual and experimental innovators are like a Kölsch beer and a vintage port. The former is best served fresh, whereas the latter ages with more pleasing mouthfeel and softening of tannins. This contrasting approach to the creative process translates to fields beyond science. Among important conceptual innovators in the modern era, Albert Einstein, Pablo Picasso, Andy Warhol, and Bob Dylan all made their landmark contributions between the ages of 24 and 36, whereas the great experimental innovators, such as Charles Darwin, Mark Twain, Paul Cézanne, Frank Lloyd Wright, and Robert Frost, made their greatest contributions between the ages of 48 and 76 [112, 113]. The nature of work is such a strong predictor of when an innovator peaks, that sometimes you don't even need data to see its effect. Take for example art. In Fig. 1.4.5, we put side-by-side works by Michelangelo and by Picasso. Can you guess who is the Kölsch beer and who is the vintage port?

Box 1.4.2 Age and adoption of new ideas: Planck's Principle

Widely held is the idea that younger researchers are more likely to make radical departures from convention. Max Planck vividly put it this way: *"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."* This school of thought argues age as an important factor driving scientific progress. In other words, if young and old scientists have different affinities for accepting new ideas, then scientific progress will have to wait for older scientists to fade in relevance, before the discipline can move forward. Yet by contrast, another school of thought argues that new ideas triumph when they are supported by more empirical evidence than the competing hypotheses. Therefore in science, reason, argument, and evidence are all that matter, suggesting that the age factors may not be as important as we thought. So is what Planck said true for science?

Although people have long suspected that younger researchers are more willing to make radical departures from convention, there has traditionally been limited evidence supporting Planck's principle. For example, Hull and colleagues found that older scientists were just as quick to accept Darwin's theory of evolution as younger scientists [114]. This finding supports the idea that scientific progress is self-correcting, guided only by truth. Yet a study focusing on 452 prominent life scientists who passed away suddenly while at their peak of their career, offers empirical support for Planck's principle [115]. After the unexpected death of a prominent scientist, her frequent collaborators—the junior researchers who coauthored papers with her—experience a sudden drop in productivity. At the same time, there is a marked increase in published work by newcomers to the field, and these contributions are disproportionately likely to be highly cited. They are also more likely to be authored by young outsiders to the field.

These results are also consistent with a study that investigates the link between age and the adoption of new ideas. By measuring new words used in a paper as a proxy for new ideas, researchers found that young researchers are much more likely than older scientists to tackle exciting, innovative topics—and when they do, the resulting work has a higher impact [111]. These results document a "Goliath's shadow" effect in science: Outsiders are reluctant to challenge the leading thinker within a field while he is still alive, but stream in and take over leadership once he passed away.

**

Regardless of whether the scientists applying for NIH funding were conceptual or experimental innovators, Elias Zerhouni was right to be concerned: One needs the funds to innovate immediately, the other to collect the evidence that would facilitate her breakthrough decades later. Overall, this chapter unveils an intimate link between age and scientific achievement, showing that a scientist's performance peaks relatively early, followed by a rapid, if brutal decline. Once we pass the peak, hope for a breakthrough starts to dim.

Or so it seemed anyway. While the research we shared here is solid, older scientists can maintain their hope. Because, as we show in the next chapter, there's more to the story.

Chapter 1.5

Random Impact Rule

Much of the research on age and achievement discussed so far has one thing in common: They focus on famous scientists, those that we laud as geniuses. So, are their conclusions relevant to us mere mortals?

This long-standing focus on prominent scientists makes sense methodologically: most of the existing knowledge in this space was tabulated by hand, scrawling down the dates of major works and estimating the scientist's age when he completed them, and sometimes locating the evidence deep in the library stacks. Equally important, information on prominent scientists is easier to come by, as they were collected in biographies and laudations.

Even today, as computers have significantly eased our ability to collect and organize data, it remains a difficult task to study individual careers, given the name disambiguation challenges discussed in Chapter 1.1 (Box 1.1.1). Yet, thanks to advances in data mining and machine learning, which uses information from research topic to the author's institution to citation patterns, the accuracy of name disambiguation has improved considerably in the past decade. As a result, career histories of individual scientists—not just geniuses, but also everyday Joes and Janes slogging it out in the field—are becoming available to researchers on a much larger scale. These advances offer new opportunities. In fact, as we will see in this chapter, the data did not just test and validate existing theories and frameworks. It upended the way we think about individual careers altogether.

1.5.1 Life Shuffled

Figure 1.5.1a shows the scientific career of Kenneth G. Wilson, the 1982 Nobel laureate in physics. We treat his first publication as the start of his academic timeline, and every time he publishes another paper, we add one more pin at the corresponding time point (academic age) in his career. The heights of each pin shows the paper's impact, approximated by the number of citations the paper received 10 years after its publication.

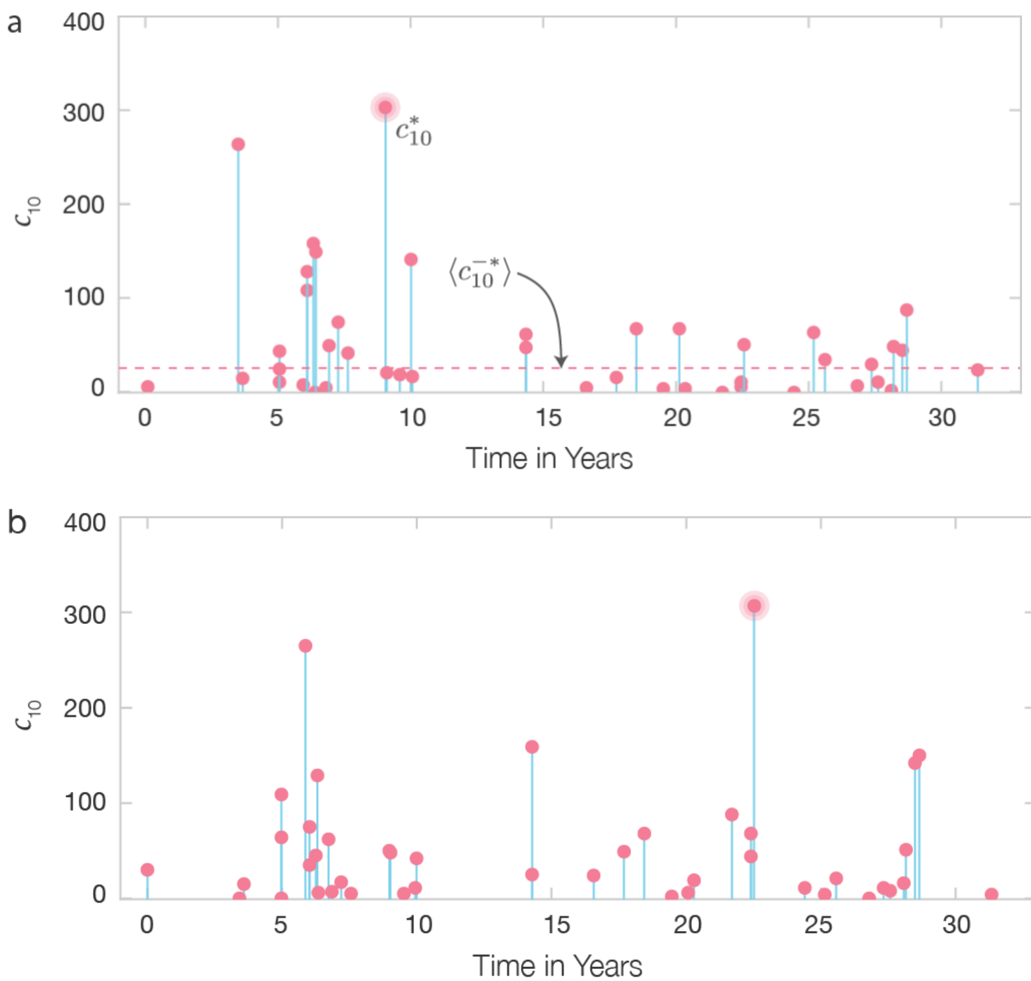

Figure 1.5.1 **Publication history of Kenneth G. Wilson**. (a) The horizontal axis indicates the number of years after Wilson's first publication and each vertical line corresponds to a publication. The height of each line corresponds to c_{10}, *i.e.* the number of citations the paper received after 10 years. Wilson's highest impact paper was published in 1974, 9 years after his first publication; it is the 17th of his 48 papers, hence t^*=9, N^*=17, N = 48. (b) Shuffled career of Wilson, where we keep the locations of the pins, but swap the impact of each paper with another one, thereby breaking the temporal ordering of when best work occurs within a career. After Sinatra *et al.* [116].

This "drop pin" view allows us to represent the career of every scientist in the same way. When repeated for tens of thousands of scientists from all kinds of disciplines, it can help us answer a simple question that has long been elusive, despite the extensive literature on geniuses: When does an *ordinary* scientist publish her best work?

Initial attempts to answer the question focused on the career histories of 2,856 physicists whose publication records span at least 20 years [116], extracted from the publication records of 236,884 physicists publishing in *Physical Review*. The dataset allowed us to identify each physicist's personal hit, i.e. the paper that collected more citations than any other paper he published. To understand when does a scientist publish his highest impact work, we measure t^*, the academic age of the scientist when the hit paper was published. Hence t^* marks, for example, the penicillium chrysogenum paper for Alexander Fleming, and the radioactivity paper for Marie Curie. But it could also denote the far less cited, yet nonetheless personal best paper by your colleague in the office next door.

Figure 1.5.2 plots $P(t^*)$, the probability that the highest impact paper is published t^* years after a scientist's first publication. The high $P(t^*)$ between 0 and 20 years indicates that most physicists publish their highest impact paper in early or mid-career, followed by a significant drop in $P(t^*)$ beyond this period. It shows that once scientists pass their mid-career, the possibility of creating breakthrough work becomes depressingly unlikely.

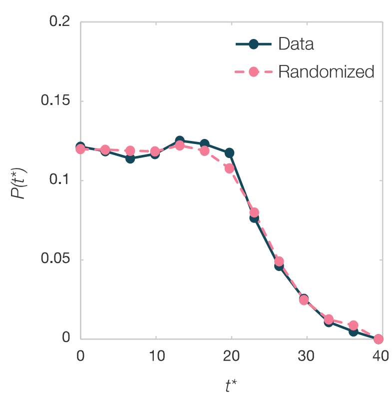

Figure 1.5.2 **The Random Impact Rule.** Distribution of the publication time t^* of the highest impact paper in scientists' career (black circles) and for randomized impact careers (black circles). The lack of differences between the two curves indicates that impact is random within a scientist's sequence of publication. After Sinatra *et al.* [116].

Yet, upon a closer examination, it turns out that the interpretation of this curve is not as straightforward as it seems at first. To see this, we next ask how would the same plot look like, if the timing of the highest impact work is driven entirely by chance?

Imagine for a second that creativity in a career is purely random. To find out how such a random career would look like, we take two pins at random and swap them, and we do this over and over, thousands of times. In doing so, we arrive at a shuffled version of each scientist's career (Figure 1.5.1b). Where does the shuffled career differ from the real one? The net productivity of the individual has not changed. The total impact of these papers didn't change either, since we did not alter the size of the pins. Nor did we change the timing of the publications. The only thing that *did* change is the order in which these papers were published. Imagine a lifetime of papers as the deck of cards you're dealt, and your highest impact paper is your ace of diamonds, then we just shuffled the order of those cards, including the ace. Your ace of diamonds now can appear anywhere—top, middle, bottom.

Next, we measured $P(t^*)$ for the shuffled careers and plotted the randomized version of $P(t^*)$ together with the real one in the same figure. To our surprise, the two curves in Fig. 1.5.2 are right on top of each other. In other words, the timing of the best works in the randomly shuffled careers is indistinguishable from the original data. What does this mean?

1.5.2 Random Impact Rule

The fact that the two distributions in Fig. 1.5.2 are the same indicates that variations in $P(t^*)$ are *fully explained by changes in productivity* throughout a career. Indeed, the randomized curve measures the variations in productivity during a scientist's career. It shows that in this sample of scientists, productivity has a peak at year 15 of a career, and it drops rapidly after year 20. This means that young scientists have a disproportionate number of breakthroughs early in their career not because youth and creativity are intertwined, but simply because they're in their most productive period. In other words, when we adjust for productivity, high impact work will occur randomly over the course of a career. We call this the *random impact rule* [116].

Staying with the card analogy, imagine you draw one card at a time from your deck, but with varying frequency. You start furiously in the beginning of your career, drawing one card after another, rapid-fire, excited by your work and hoping to find that ace. This rapid-fire period is then followed by a gradual decline, where you slow down how frequently you reach out to the deck. Now, if the deck was well-shuffled beforehand, and you draw a lot more cards during the first 20 years than during your later period, when will you most likely encounter the ace? During the first 20 years, of course. In other words, the first two decades of your career are not more creative than the later 20 years. You draw an ace early in your career simply because you try harder.

To more directly test the random impact rule, we can look at where the ace appears within the deck. For that we calculate the position of the highest impact paper N^* in the sequence of N publications of a scientist. Then, we measure $P(N^*/N)$, i.e. the probability that the most cited work is published early (small N^*/N) or late ($N^*/N \approx 1$) in the sequence. If the random impact rule holds true, $P(N^*/N)$ should follow a uniform distribution: That is, with equal probability we should find the hit paper at any positions of N^*/N. In technical terms, this means that the cumulative distribution $P^>(N^*/N)$ must decrease as a straight line, following $(N^*/N)^{-1}$. As Figure 1.5.3a shows, the data follows exactly what the random impact rule predicts.

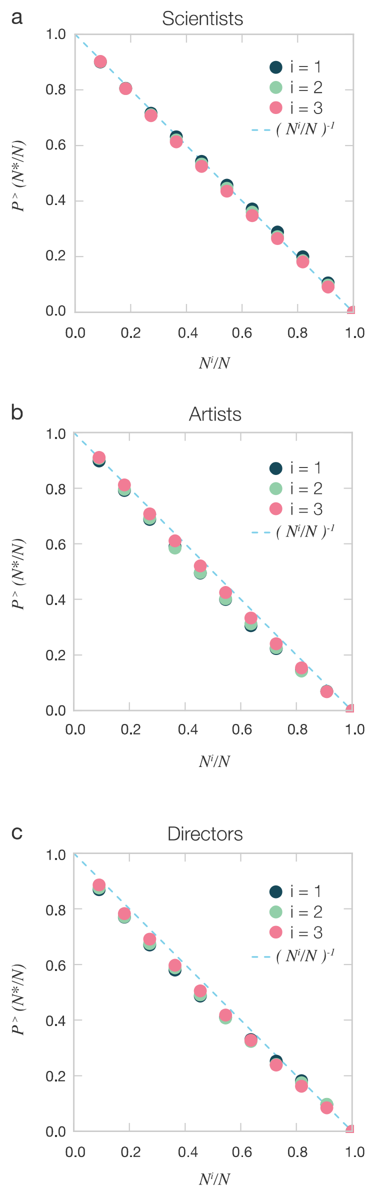

Figure 1.5.3 **Random Impact Rules Across Creative Domains.** Cumulative distribution $P^>(N^*/N)$, where N^*/N denotes the order N^* of the highest impact paper in a career, varying between $1/N$ and 1. The cumulative distribution of N^*/N is a straight line with slope -1, indicating that N^* has the same probability to occur anywhere in the sequence of works one produces. Figure shows $P^>(N^*/N)$ for careers of 20,040 scientists (a), 3,480 artists (b), and 6,233 film directors (c) [112]. For each creative individual, we identified her top three highest impact works (publications, artworks, movies) and measured their relative position within her career (citations, auction prices, ratings recorded in the *Internet Movie Database* (*IMDb*). The panels demonstrate that, across all three careers, the timing of each of the three highest impact work is random within the sequence of works one produces. After Liu *et al.* [117].

But why stop at the highest impact paper? What about the second highest? How about the third? As you may have guessed, the same pattern emerges (Fig. 1.5.3a). The cumulative distribution follows a clear straight line. That is, the big break of your career can come at any time, and this pattern is not limited to your highest impact work—your other important works are equally random [117]. And this random impact rule holds not only in scientific careers, but also in careers across different creative domains, like artists and film directors [117] (Fig. 1.5.3).

The random impact rule has deep roots in the literature, dating back to the 1970s work by Simonton, who proposed the "constant-probability-of-success" model [2, 118-121]. Researchers have long suspected that the same rule holds for the arts, such as literary and musical creativity [118]. But it took more than 40 years to collect the appropriate datasets to have it formally tested.

Box 1.5.3 Lost Winners.

The results of Chapter 1.4 showed that the prize-winning works by Nobel laureates tend to occur early within their career. By contrast, this chapter shows that ordinary scientific careers are governed by the random impact rule. Does the random impact rule apply to Nobel careers as well [122]? To find out we measured the positions of the prize-winning work and most-cited work within the sequence of papers published before being awarded the Nobel Prize (51.74% of the most-cited papers were also the prize-winning papers), finding that both tend to occur early within the sequence of papers (Figure B1.5.3). This suggests that compared with ordinary scientists, Nobel laureates tend to do their best work disproportionately early in their careers.

Yet, there is a selection effect we must confront---since the Nobel prize in science has never been awarded posthumously, those who produced ground-breaking works early were more likely to be recognized. To test this hypothesis, we removed prize-winning papers, which are subject to this selection bias, and measured the timing of each of the three remaining highest impact works for Nobel laureates, finding that they are all distributed randomly within their career (Fig. B1.5.3). This means, apart from the prize-winning work, all other important works in a Nobel-winner career follow the random impact rule. One implication of this selection bias is the existence of 'lost winners,' scientists whose deserving works were not recognized by a Nobel simply because their work came late in their career. This finding may be especially alarming given that the time lag between discovery and recognition has been growing (see Box 1.4.1).

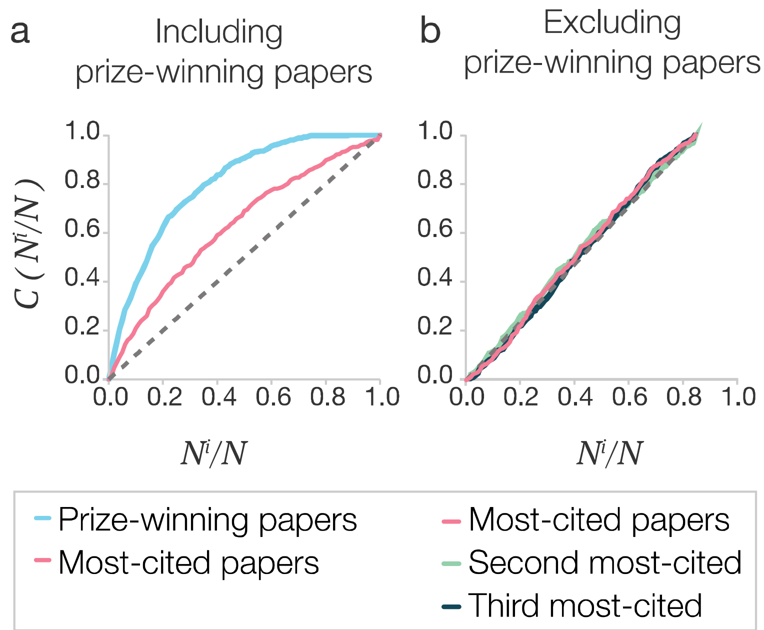

Figure B1.5.3 **Career patterns of Nobel laureates.** a. The cumulative distributions of relative positions (N_i/N) of the prize-winning papers and the most-cited papers within the sequence of all papers before being awarded the Nobel prize. The dashed line indicates the predictions of the random impact rule. b. To eliminate potential selection bias in the timing of the prize-winning work, we removed prize-winning papers and calculated the relative position of the top three most-cited papers among all the papers published before the award, finding that these papers follow the random impact rule. After Li *et al* [122].

The random impact rule changes our understanding of when breakthroughs happen in an individual's career. Indeed, decades of research has documented that major discoveries often come early in a scientist's career. This has led to the view that creativity equals youth, a myth deeply ingrained in the popular culture. The random impact rule helps us decouple age and creativity. It tells us that the chance of a breakthrough is completely random within the sequence of works produced in a career. To be precise, every project we do has the same chance of becoming our personal best. What is not random, however, is productivity: younger researchers are more eager to try over and over, putting papers out one after another. If the impact is random within an individual's range of projects, then it is inevitable that, statistically speaking, it will happen early on in an individual's career, when productivity is high.

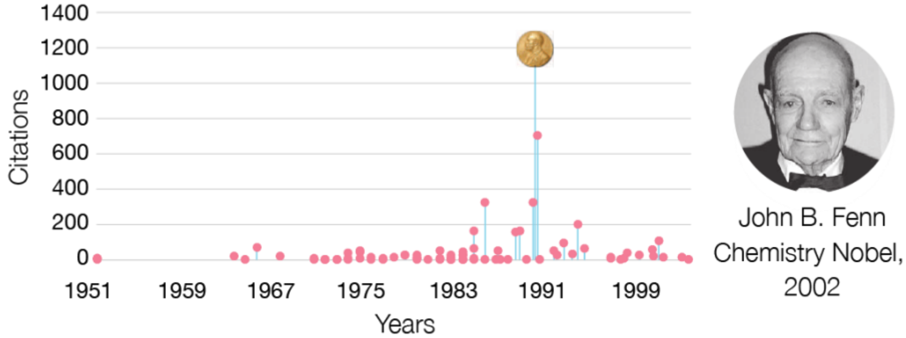

Figure 1.5.4 **The academic career of John B. Fenn, the 2002 Nobel laureate in Chemistry.**

The random impact rule thus offers a new perspective on the role of productivity: it tells us that trying over and over is important for making that long-awaited breakthrough. Indeed, for those that keep trying, the breakthrough may not be as elusive. A wonderful example is offered by John Fenn. He identified a novel electrospray ion source at the very end of his official academic career, just as he was forcefully retired by Yale University. Undeterred, he left Yale, took a new faculty position at Virginia Commonwealth University, and continued his investigation, which eventually led to the discovery of electrospray ionization, for which, 15 years later, he received his Nobel prize. Overall, his example, together with the random impact rule, tells us that for those who do not let their productivity drop in the later part of their career, impact does not wane.

While the random impact rule deepens our understanding of patterns underlying scientific careers, it also raises a new puzzle: If the timing of the biggest hit is random, what is not random in a career?

Box 1.5.4 The old myth of the young entrepreneur

The youth=creative dogma is not limited to science—it is deeply ingrained in the world of entrepreneurship as well. Indeed, the average age of the winner of the TechCrunch awards in Silicon Valley is 31, and those named "Top Entrepreneurs" by Inc. and Entrepreneur Magazines have an average age of 29. Similarly, investment money also flows to young entrepreneurs: the average age of the founders backed by one of the best-known VC firms, Sequoia, is 33, and those supported by Matrix Ventures were on average 36. But does youth imply success in Silicon Valley?

A recent analysis shows otherwise. By combining tax-filing data, U.S. Census information, and other federal datasets, researchers compiled a list of 2.7 million company founders [123]. Their analysis revealed that, contrary to popular thinking, the best entrepreneurs tend to be middle-aged. Among the very fastest-growing new tech companies, the average founder was 45 at the time of founding. Furthermore, a 50-year-old entrepreneur is nearly twice as likely to have a runaway success as a 30-year-old.

These results show that entrepreneurial performance rises sharply with age. Indeed, if you were faced with two entrepreneurs and knew nothing about them besides their age, contrary to the prevailing wisdom, you would do better, on average, by betting on the older one.

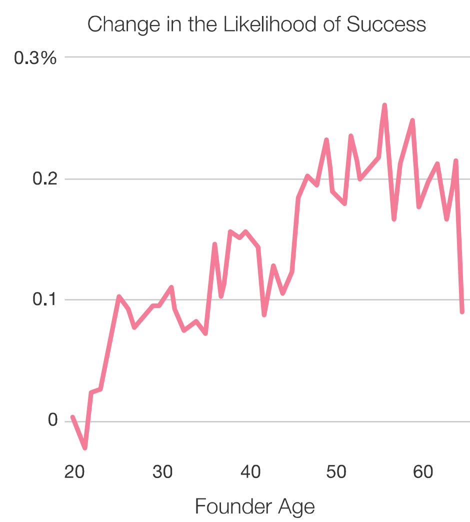

Figure B1.5.4 **Older Entrepreneurs are more likely to succeed.** The probability of startup success rises with age, at least until the late 50s. The y-axis represents the OLS (ordinary least square) regression coefficient for age variables, measuring change in the likelihood of extreme success relative to a 20-year-old founder. Here "extreme startup success" is defined as the top 0.1% of startups in employment growth over five years [123]. After Azoulay et al. [124].

Chapter 1.6

The *Q* Factor

What distinguishes successful scientists from their peers? Our ability to identify the roots of a successful career is limited by the fact that productivity, impact, and luck are intimately intertwined. Indeed, if breakthroughs in a career occur at random, what is the relevance of luck, talent, or hard work? Can we at all separate the innate talent and ability of a scientist from his luck? To understand these questions, let's start with a thought experiment: How likely is it that Einstein's exceptional output may emerge purely by chance?

1.6.1 Pure coincidence

Given an infinite length of time, a chimpanzee punching at random on a typewriter will surely type out a Shakespeare play. So, with a sufficiently large number of scientists in the world, shouldn't we expect that chance alone will inevitably lead to someone as impactful as Einstein?

The random impact rule discussed in Ch. 1.5 allows us to build a "null" model of scientific careers to answer this question. Let us assume for the moment that for a scientist a publication is the equivalent of drawing a lottery ticket. In other words, we assume that talent plays no role, allowing us to ask, how does a career driven by chance alone look like?

In a random career the impact of each paper is determined solely by luck. This means that we simply pick a random number from a given impact distribution and assign it to the paper published by our scientist.

This procedure allows us to generate a collection of "synthetic" careers which are purely driven by chance. For convenience, we will call this procedure the random model, or the *R*-model.

In some ways, these random careers may appear quite similar to the real ones. For example, they will show individual differences in career impact, as some scientists are inevitably luckier than others when they select their random numbers. And each career will obey the random impact rule as well: Since the impact of every paper is chosen randomly, the highest-impact work will be random in the sequence of papers published by each scientist. But will these random scientists differ from the real ones?

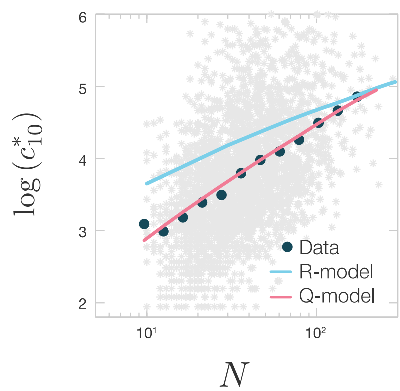

Figure 1.6.1 **Scientific careers are not random**. Scatter plot shows the citation of the highest impact paper, c_{10}^* vs the number of publications N during a scientist's career. Each grey dot corresponds to a scientist. The circles are the logarithmic binning of the scattered data. The cyan curve represents the prediction of the *R*-model, which shows systematic deviation from data. The red curve corresponds to the analytical prediction of the *Q* model. After Sinatra *et al.* [116].

If each paper's impact is drawn randomly from the same impact distribution, a more productive scientist will draw more tickets, hence will more likely stumble upon a high impact publication. In other words, the *R*-model predicts that more productive scientists are more likely to make breakthroughs. To test this effect, we measured how the impact of a scientist's most-cited paper, $\langle c_{10}^* \rangle$, depends on her productivity, N. We find that indeed, the more papers a scientist publishes, the higher is the impact of her highest impact paper. Yet, not high enough: The measurements indicate that in real careers the impact of the highest impact work increases faster with N than the *R*-model predicts (Fig. 1.6.1). In other words, as scientists become more productive, their home run papers are much more impactful than what we would expect if impact is like a lottery ticket, randomly assigned. This indicates that something is missing from our null model. It's not hard to guess what: scientists inherently differ from each other, in either talent or ability or other characteristics relevant to producing high impact works. This suggests that highly productive scientists don't stand out in productivity alone---they possess something that the low productivity scientists do not, which helps them publish higher impact work. Next, we adjust the model to account for the fact that not all scientists are alike.

1.6.2 The *Q*-model

Each scientific project starts with an idea. A light bulb goes off, prompting the scientist to think, "I'm curious if that idea might work." But it is hard to evaluate the inherent importance or novelty of the idea in advance. Not knowing what its true value is, let's assume that the idea has some random value r. Some ideas are of incremental importance, only interesting to a limited number of people in our immediate field, which would correspond to a modest r. Occasionally, though, we stumble upon an idea that could be transformative—that is if it works out. The better the idea—the larger its r value—the more likely it is to have high impact.

But having a good idea isn't enough. The ultimate impact of the project depends on a scientist's ability to turn that idea into a truly impactful product. One may begin with a terrific idea, but lacking the necessary expertise, experience, resources, or thoroughness to bring out its full potential, the end result will suffer. Yet, turning that idea into a discovery requires an ability that varies from person to person. So, let us assign a parameter Q to characterize an individual's ability to turn a random idea r into a discovery of a given impact.

In other words, let us assume that the impact of each paper we publish, c_{10}, is determined by two factors: luck (r) and the Q_i parameter unique to individual i. The combination of the two could invoke some complicated functions. To keep it simple, we assume a simple linear function, writing

$$c_{10} = rQ_i. \tag{1.6.1}$$

Equation (1.6.1) has multiple assumptions behind it:

(a) When we start a new project, we pick a random idea r from a pool of possibilities. Scientist may pick their r from the same distribution $P(r)$, as we all have access to the same literature hence the same pool of knowledge. Or each scientist may pick her r from her own $P(r)$, as some scientists are better at picking good ideas than others.

(b) Scientists differ in their Q parameter, which characterizes the ability of an individual to take the same idea but turn it into works with variable impacts. If a scientist with a low Q-factor comes across an idea with huge r value, despite the idea's inherent potential, the project's impact will be mediocre, since the resulting product rQ is diminished by the scientist's limited Q. On the other hand, a high Q scientist may also put out weak or mediocre works, if the idea he started with was poor (small r). Truly high-impact papers are those perfect-storm instances when a high Q individual stumbles upon a brilliant idea (high r). In other words, the model assumes that the ultimate impact of a paper is the product of two factors: a yet-to-be-realized potential of an idea and a scientist's ability to realize it.

(c) Finally, productivity matters: Scientists with high N, even with the same Q and $P(r)$, have more chances to stumble across a high r project, turning it into a paper with a high impact (c_{10}).

The problem is that none of these factors are expected to be independent of each other: high Q individuals may also have a talent at recognizing high potential projects, hence their $P(r)$ distribution could be skewed towards higher r values. And those who publish higher impact papers may also have more resources to publish more papers, hence will have higher productivity. In other words, the outcome of the model (1.6.1) is determined by the joint probability $P(r, Q, N)$, with unknown correlations between r, Q, and N. To understand what real careers look like, one needs to measure the correlations between the three parameters. This measurement led to the covariance matrix [116],

$$\Sigma = \begin{pmatrix} \sigma_r^2 & \sigma_{r,Q} & \sigma_{r,N} \\ \sigma_{r,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{r,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} 0.93 & 0.00 & 0.00 \\ 0.00 & 0.21 & 0.09 \\ 0.00 & 0.09 & 0.33 \end{pmatrix}, \tag{1.6.2}$$

which makes two unexpected predictions about individual careers:

i. $\sigma_{r,N} = \sigma_{r,Q} \cong 0$ indicates that the value of an initial idea (r) is largely independent of a scientist's productivity N or her Q-factor. Therefore, scientists source their ideas randomly from a $P(r)$ distribution that is the same for all individuals, capturing a universal—that is, scientist-independent—luck component behind impact.

ii. The nonzero $\sigma_{Q,N}$ indicates that the hidden parameter Q and productivity N do correlate with each other, but its small value also shows that high Q is only slightly associated with higher productivity.

The lack of correlations between the idea value r and (Q, N) allows us to analytically calculate how the highest-impact paper c_{10}^{*} is expected to change with productivity N. As shown in Fig. 1.6.1, the prediction of the Q-model is now in excellent agreement with the data, indicating that the hidden Q-factor and variations in the productivity N can explain the empirically observed impact differences between scientists, correcting the shortcomings of the R-model.

The fact that we need other parameters besides luck to characterize a scientist's career impact makes sense. It's not hard to imagine that there are differences between individual scientists, which need to be accounted for to have an accurate description of real careers. What is surprising, however, is that we appear to need only *one* additional parameter besides luck. Incorporating the Q factor *alone* seems sufficient to explain what distinguishes one scientist from the other.

What exactly does the Q-model offer that the R-model misses? The R-model's failure tells us that a successful career isn't built on chance alone. Indeed, the Q factor pinpoints a critical characteristic of a career: Great scientists are *consistently* great across their projects. True, each scientist probably has one important paper they are known for. But that paper didn't appear by chance. A great scientist's second best paper, or her third, or, for that matter, *many of* her other papers are unusually highly cited as well. Which means that there must be something unique about a researcher who can consistently produce outstanding papers. That unique characteristic is what Q captures. In other words, while luck is important, by itself it won't get you far. It is the Q-factor that turns luck into a consistently high-impact career.

1.6.3 What is your *Q*?

The Q-model not only helps us break down the elements associated with high impact careers, but it also allows us to calculate the Q factor for each scientist, based on their sequence of publications. The precise solution for Q is somewhat mathematically involved, but once a scientist has published a sufficient number of papers, we can approximate her Q by a rather simple formula [116].Consider a career in which each paper j published by scientist i collects $c_{10,ij}$ citations in ten years. Start by calculating $\log c_{10,ij}$, for each paper, then averaging over all logarithmic citations. The Q_i is the exponential of that average,

$$Q_i = e^{\langle \log c_{10,i} \rangle - \mu_p}, \tag{1.6.3}$$

where μ_p is a normalization factor, that depends on the career output of all scientists. Given its dependence on the average of $\log c_{10,ij}$, Q is not dominated by a single high (or low) impact discovery, but captures, instead, a scientist's *sustained* ability to systematically turn her projects into high (or low) impact publications. To better understand Q, let's look at an example.

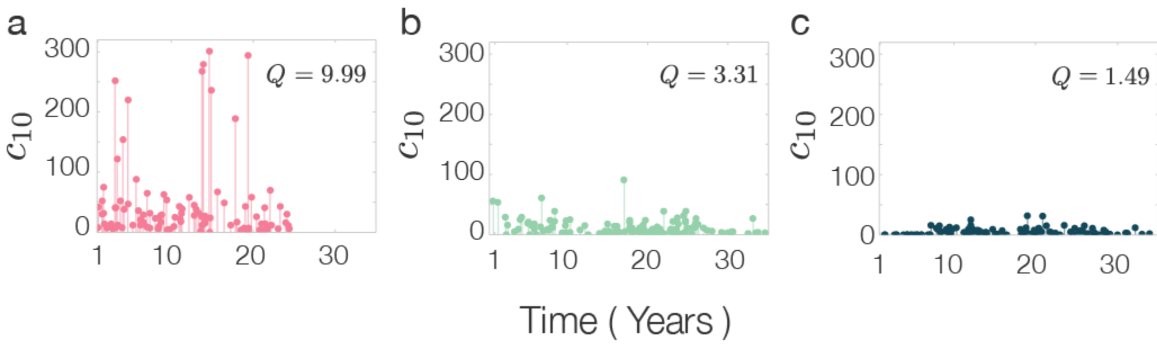

Figure 1.6.2 **Careers with different *Q*-factor**. The figures illustrate career histories of three scientists with comparable productivity ($N \simeq 100$), documenting the notable differences between the papers published by them, given their different Q-factors.

The three scientists shown in Fig. 1.6.2 have comparable productivity, all having published $N \simeq 100$ papers. Yet, their career has noticeably different impact. Using (1.6.2), we can calculate the Q-factor for each of them, obtaining Q=9.99, 3.31 and 1.49, respectively. As Fig. 1.6.2 illustrates, the Q-factor captures persistent differences in impact across a scientist's sequence of publications: the Q=9.99 researcher produces one high-impact papers after another. In contrast, the work by the Q=1.49 scientist garners consistently limited impact. The one in the middle may occasionally publish a paper that jumps out of the pack—that is, if he gets lucky—but its impact is dwarfed by what the researcher to his left has managed to achieve. Therefore, Q describes scientists' different ability to take random projects r and systematically turn them into high (or low) impact publications. Any single one of the projects may be influenced by luck. But inspecting project after project, the true Q of a scientist starts to stand out.

There are many benefits of the mathematical equation (1.6.3). To begin with, it allows us to estimate expected impact of a career. For example, how many papers does a scientist need to publish before he can expect one of them to reach a certain level of impact? According to (1.6.3), a scientist with the somewhat low Q=1.2, similar to that shown in Fig. 1.2c, needs to write at least 100 papers if he wants one of them to acquire 30 citations over a 10-year period. On the other hand, an equally productive scientist with Q=10 can expect to have at least one of her papers reach 250 citations over the same 10-year period.

Consider next what happens if both scientists ramp up their productivity. Regardless of Q, a higher productivity increases the chance of stumbling upon a fabulous idea, i.e. picking a higher r value. Hence, we expect the highest-impact paper to increase for both careers. If the low-Q scientist doubles his productivity, though, he'll only enhance the impact of his best paper by seven citations. Compare that with the high-Q scientist, who will see a boost of more than 50. In other words, for a scientist with limited Q, an increase in productivity doesn't substantially improve his chances for a breakthrough—simply working harder isn't enough.

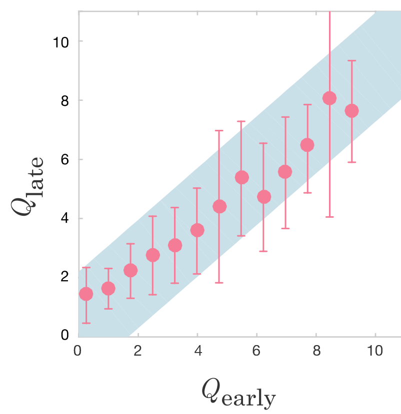

Figure 1.6.3 **The Q-factor appears relatively stable during a career**. We compare the Q parameter at early-career (Q_{early}) and late-career (Q_{late}) stage of 823 scientists with at least 50 papers. We measured the two values of the Q parameters using only the first and second half of published papers, respectively. We perform these measurements on the real data (circles) and on randomized careers, where the order of papers is shuffled (gray shaded areas). For most of the careers (95.1%), the changes between early- and late-career stages fall within the fluctuations predicted by the shuffled careers, suggesting that the Q parameter is relatively stable throughout a career.

Does the Q-factor increase with age and experience? As a scientist, we would like to think that as we advance in our careers, we become better at translating our ideas into high-impact publications. To test the stability of the Q parameter throughout the overall career, we consider careers with at least 50 papers and calculate their early and late Q parameters (Q_{early} and Q_{late}, respectively) using Eq. 1.6.3 on the first and second half of their papers, respectively. As Figure 1.6.3 shows, Q_{late} is proportional to Q_{early}, indicating that the Q parameter does not systematically increase or decrease over a career. In other words, a scientist's Q-factor appears relatively stable during her career. This raises a tantalizing question: can the Q parameter predict the career impact of a scientist?

1.6.4 Predicting impact

To understand if the Q-factor is more effective at gauging the overall impact of an individual's work, we throw several metrics we've discussed so far in this book into a kind of horse race. We start by checking how well these different measures can forecast Nobel Prize winners [116]. For that we can rank physicists based on their productivity N, total number of citations C, citations of the highest-impact paper c_{10}^{*}, h-index, and Q. To compare the performance of each ranking, we use a ROC-plot that measures the fraction of Nobel Laureates at the top of the ranked list. Figure 1.6.4 shows that, overall, cumulative citation measures, like the number of citations collected by the highest impact paper of a scientist, and the total career citations, do reasonably well. And the h-index is indeed more effective than these citation measures, ranking Nobel-winning careers at a higher accuracy. The worst predictor, interestingly, is the productivity, representing the number of papers published by scientists. In other words, simply publishing a lot is not a path toward the Nobel.

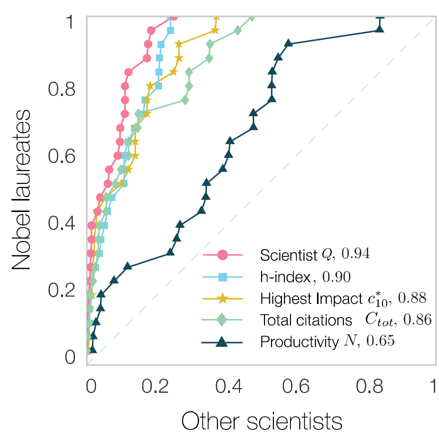

Figure 1.6.4 **Predicting Nobel Laureates**. ROC plot captures the ranking of scientists based on Q, productivity N, total number of citations C, citations of the highest-impact paper c_{10}^{*} and h-index. Each curve represents the fraction of Nobel laureates versus the fraction of other scientists for a given rank threshold. The diagonal (no-discrimination line) corresponds to random ranking; the area under each curve measures the accuracy to rank Nobel laureates (reported in the legend, with 1 being the maximum). After Sinatra *et al.* [116].

Yet, despite the strength of the h index, and other citation-based measures, the career-based Q factor wins out, predicting Nobel careers more accurately than all other measures tested in Fig. 1.6.4. Therefore, while the h index remains an excellent measure for the overall impact of a scientist, Q factor seems to offer even better predictive power. What does Q capture that h misses?

To understand the difference between Q and h, let's look at a randomized field experiment exploring how economists assess each other's CVs [125]. Professors from 44 universities were randomly selected from among the top 10% research-based universities in the world. These professors were asked to rate fake CVs based on the publication histories outlined in those CVs. The fake CVs contained either a publication history that listed only papers that had appeared in elite journals, or a publication history with *both* the elite papers and additional publications that had appeared in lower-tier journals. Economics is an excellent discipline for such a test, because there is a rather stable hierarchy of prestige among the journals. In other words, every economist tends to agree which journals are higher- and which are lower-tier. So, would including publications from well-known, respected, but lower-ranked journals alongside papers from higher-ranked ones make a CV stronger? If so, how much stronger?

The survey respondents were asked to rate the CVs on a scale of 1 to 10, with 1 being the absolute worst and 10 the absolute best. On average, the short CVs listing only top publications received a rating of 8.1. The long CVs—which, remember, contained the *same* top-tier papers as the short CVs but with additional lower-tiered publications—received an average rating of 7.6. This means, a short CV was systematically preferred to a long CV, even though they both featured identical elite publication histories. In other words, the additional publications from lower-tier journals not only didn't help, they actually negatively affected the experts' assessment.

These results are both intuitive and puzzling. On the one hand, they lend evidence to the often-held conjecture that publishing too frequently in lower-ranked journals may work against you. On the other hand, though, these results simply don't make sense if we think in terms of the metrics we use to assess scientific achievement. Indeed, of all the measures we've discussed, from productivity to total number of citations to *h* index, each increases monotonically with the number of papers. From this perspective, it will always make sense to publish another paper, even if its impact is limited. First, it will certainly increase your publication count. Second, even if the paper wasn't published in a prestigious journal, with time it may accumulate citations that will contribute to your overall impact, and perhaps even to your *h* index. At the very least, it can't hurt.

Except these results show us that it *does* hurt. And that makes sense if we think in terms of the *Q* factor. In contrast to other measures, *Q* doesn't simply grow with an expanding publication list. Instead, it depends on if the additional papers are, on average, better or worse than your other papers. In other words, the purpose of *Q* is to quantify a scientist's consistent ability to put out high impact papers over the course of her career. It takes into account all papers, not just the high-impact ones. Hence, if you already have a stellar publication record and publish a few more papers, these additional papers will enhance each of your other metrics, but they aren't guaranteed to increase your Q factor. In fact, unless your new papers are on par with the papers you typically publish, they may actually lower it.

The superior accuracy of the Q factor in predicting career impact therefore illustrates that for a career, consistency matters. This conclusion, together with the stability of the Q parameter suggested by Fig. 1.6.3, paints a picture of rather monotonic careers: we all start our careers with a given Q, high or low. That Q-factor determines the impact of each paper we publish and stays with us until retirement. But is it true? can we ever break this robotic monotonicity? In other words, do we ever experience periods in our career where we are actually really good at what we do?

Chapter 1.7

Hot Streaks

In physics, 1905 is known as the “miracle year”. It is the year that Einstein published four discoveries that changed the discipline forever. By the summer, he’d explained Brownian motion, which pointed to the existence of atoms and molecules; discovered the photoelectric effect, for which he was awarded the Nobel Prize fifteen years later, representing a pivotal step toward quantum mechanics; and developed the theory of special relativity, which changed the way we think about time and space altogether. Then, before the year ended, he scribbled down the world’s most famous equation: $E=mc^2$.

Einstein’s 1905 can be described as a “hot streak,” a burst of outstanding achievements. Hot streak is a familiar term in sports or gambling. If in basketball you land two shots without even touching the rim, you are not at all surprised when you land that third one. You probably could even “feel” that strange and magical momentum. While for decades hot streaks were taken as a given, in 1985 a much-quoted psychology study contested their existence [126], concluding that basketball is no streakier than a coin toss. The finding suggested that the concept is a fallacy, rooted in our psychological biases for small numbers—it only *feels* that you’re on a roll, yet in reality that streak of luck is what one would expect by chance. The 1985 study triggered a still ongoing debate among psychologists and statisticians [127-130], with the latest statistics papers arguing that the fallacy may itself be a fallacy, and hot streaks do seem to exist in sports after all [127, 131]. Yet, as the debate unfolds in sports, gambling, and financial markets, it raises an intriguing question: If we define our career as the sequence of papers a scientist produces over a lifetime, do we ever experience hot streaks in our career?

1.7.1 Bursts of hits

Across the careers of scientists, artists, and film directors, we've seen that the three biggest hits of a career each occur at random within the sequence of works one produces. This finding tells us that creativity is random and unpredictable, with chance playing an outsized role in the timing of key achievements. In other words, scientific careers are unlikely to include hot streaks. Yet, at the same time, the random impact rule raises a puzzling question: What happens after we finally produce a breakthrough?

The Matthew effect tells us that winning begets more winnings. So, if we produce one big hit, even if its timing is random, it would follow that we'd produce more hits afterwards. Yet according to the random impact rule, the opposite seems to be true. Indeed, if the impact of each work in a career is truly random, then one's next work after a hit may be more mediocre than spectacular, reflecting regression toward the mean. So, are we really regressing toward mediocrity after we break through?

To answer this question, we examined the *relative* timing of hit works within a career [117]. Specifically, we asked: Given when someone produced their best work, when would their second-best work be? To measure the correlation between the timing of the two biggest hits within a career (e.g., N^* and N^{**}) we calculate the joint probability $P(N^*, N^{**})$ for the two of them to occur together, and compare it with a null hypothesis in which N^* and N^{**} each occur at random on their own. Mathematically, this is captured by the normalized joint probability, $\varphi(N^*, N^{**})=P(N^*, N^{**})/P(N^*)P(N^{**})$, which is best represented as a matrix (Fig. 1.7.1). If $\varphi(N^*, N^{**})$ is approximately 1, then the chance to find the biggest and the second biggest hit to occur together within a career is about what we would expect if both occur at random. If, however, $\varphi(N^*, N^{**})$ is above 1, it means that N^* and N^{**} are more likely to appear together, corresponding to a correlation that is not anticipated by the random impact rule.

Figure 1.7.1 shows $\varphi(N^*, N^{**})$ for careers across sciences, arts, and movies, leading to three important conclusions:

- First, $\varphi(N^*, N^{**})$ is notably higher along the diagonal elements of matrices, indicating that N^* and N^{**} are much more likely to colocate with each other than expected by chance. In other words, if we know where your biggest hit happens in the course of your career, we would know quite well where your second biggest hit is—it will be just around the corner of your biggest hit. The two most important works of a scientist are on average 1.57 times more likely to occur back-to-back than we'd expect to see by chance.

- Second, φ features a roughly even split across the diagonal, which means that there is a comparable likelihood of the biggest hit arriving either before or after the second biggest hit.
- Third, the colocation pattern is not limited to the two highest impact works within a career. Indeed, if we repeat our analyses for other pairs of hits, such as N^{*} vs. N^{***} and N^{**} vs. N^{***}, we find the same pattern. It is not only the top two hits which appear close to each other. The *third* hit is also in the vicinity of the first two.

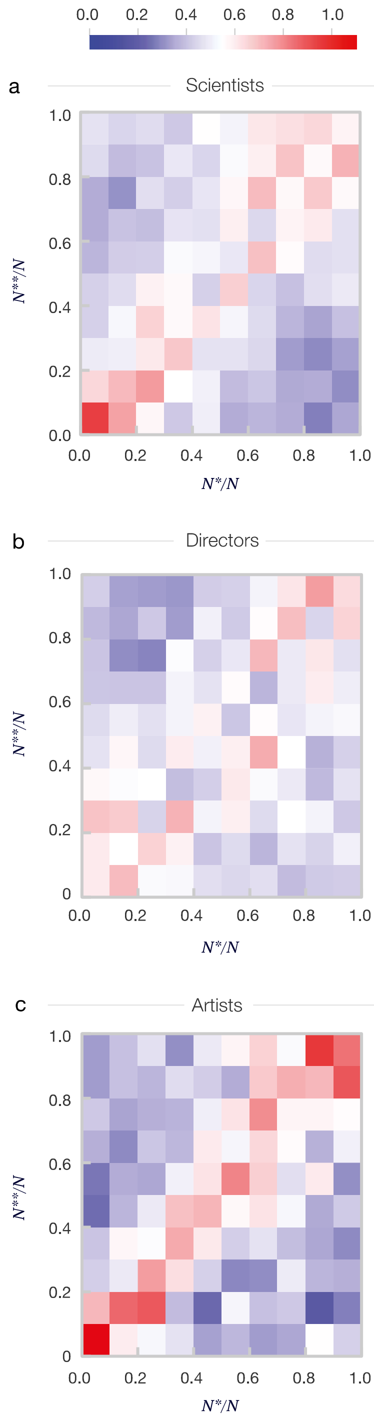

Figure 1.7.1 **Hot-streaks in artistic, cultural and scientific careers**. $\varphi(N^{*},N^{**})$, color-coded, measures the joint probability of the top two highest-impact works within the career of a) scientists, b) artists, and c) film directors. $\varphi(N^{*},N^{**})> 1$ (red along diagonal) indicates that two hits are more likely to occur next to each other than expected by chance. After Liu *et al.* [117].

These results offer a more nuanced portrait of individual careers. Indeed, while the timing of each highest impact works is random, *the relative timing of the top papers in an individual's career* follows highly predictable patterns. As such, individual career trajectories are not truly random. Rather, hit works within a career show a high degree of temporal clustering, with each career being characterized by bursts of high-impact works coming in close succession. Further, these patterns are not limited to scientific careers. The hits of artists and film directors show similar temporal patterns, suggesting that there's a ubiquitous clustering of successful work emerging across a range of creative careers. What are the mechanisms responsible for these remarkable patterns?

1.7.2 The hot-streak model

To understand the origin of the observed bursts of career hits, let's start with the Q model discussed in the previous chapter. Imagine that every time a scientist publishes a paper, the work's impact is determined by a random number generated from a given distribution, which is fixed for that individual. Since citations are well approximated by lognormal distributions, let's assume for convenience that their logarithm is drawn from a normal distribution, with average Γ_i. A career generated by this null model will nicely follow the random impact rule: each hit, including the biggest one, occurs randomly within a career [2, 116]. It all depends on luck—the pay-off of a lottery ticket drawn from a distribution of random numbers.

However, this null model fails to capture the temporal correlations documented in Fig. 1.7.1. The main reason is shown in Fig. 1.7.2a–c, where, for illustration purposes, we selected one individual from each of the three data sets and measured the dynamics of Γ_i during his or her career. These examples indicate that Γ_i is not constant. Rather, it is characterized by a baseline performance (Γ_0) until a certain point in a career, after which it is elevated to a higher value Γ_H ($\Gamma_H > \Gamma_0$). That elevated performance is sustained for some time before eventually falling back to a level similar to Γ_0. This observation raises an interesting question: Could a simple model, which assumes everyone experiences a short period like Γ_H, explain the patterns documented in Fig. 1.7.1?

Remember, this new model introduces only a simple variation over the Q-model—a brief period of elevated impact. But interestingly, that small change can account for the empirical observations that the random impact rule or the Q model failed to capture (Fig. 1.7.1). During the period in which Γ_H operates, the individual seemingly performs at a higher level than her typical performance (Γ_0). We call this model the *hot-streak model*, where the Γ_H period corresponds to the hot streak.

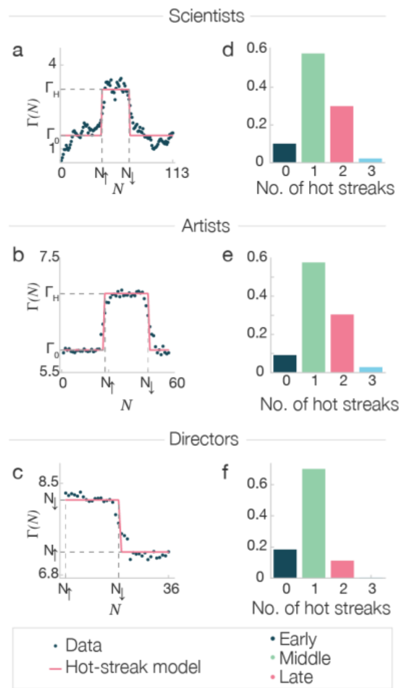

Figure 1.7.2 **The hot-streak model.** a–c, Γ(N) for one scientist (a), artist (b) and film director (c), selected for illustration purposes. It shows that real careers can have strong temporal variations, characterized by brief period of elevated impact. d–f, Histogram of the number of hot streaks in a career shows that hot streaks are ubiquitous yet usually unique. Most people have hot streaks, and they mostly likely have just one. After Liu *et al*. 2018 [117].

The true value of the hot-streak model emerges when we apply it to career data, allowing us to obtain the parameters which characterize the timing and the magnitude of hot streaks for individual scientists, leading to several important observations:

1. The measurements indicate that hot streaks are ubiquitous across creative careers. About 90% of scientists experience at least one hot streak, as do 91% of artists and 82% of film directors. This means that hot streaks are not limited to Einstein's miracle year of 1905. The director Peter Jackson's hot streak lasted for 3 years, while he directed *The Lord of the Rings* series, and Vincent van Gogh's occurred in 1888, the year he moved from Paris to Southern France, producing works like *The Yellow House*, *Van Gogh's Chair*, *Bedroom in Arles*, *The Night Café*, *Starry Night Over the Rhone*, and *Still Life: Vase with Twelve Sunflowers*.

2. Despite their ubiquity, hot streaks are usually unique in a creative career. Indeed, when we relaxed our algorithm to allow for up to three hot streaks, we found that 68% of high impact scientists experience only one hot streak (Figs. 1.7.2d). A second hot streak may occur but is less likely, and occurrences of more than two are rare.
3. Hot streaks occur randomly within a career, offering a further explanation for the random impact rule. Indeed, if the hot streak occurs randomly within the sequence of works produced in a career, and the highest impact works are statistically more likely to appear within the hot streak, then the timing of the highest impact works would also appear to be random.
4. Hot streaks last for a relatively short period. The duration distribution of hot streaks peaks around 3.7 years for scientists, indicating that, whether it happens early or late in a scientist's career, it most likely lasts for only four years. For artists or film directors, it lasts about five years.
5. While people create more impactful work during hot streaks, they aren't more productive than we would expect during that time. It's just that what they do produce is substantially *better* relative to the rest of their work.

Why do we have hot streaks in the first place? There are several plausible hypotheses. For example, innovators may stumble upon a groundbreaking or temporally resonant idea, which then results in multiple high-impact publications. Alternatively, since high-impact outcomes are increasingly produced by teams, an individual's hot streak may reflect a fruitful, repeated, but short-lasting collaboration. Or perhaps it is related to shifting institutional factors like tenure track, corresponding to career opportunities that augment impact but last for a fixed duration. Analyses of real careers suggest that, while plausible, none of these hypotheses alone can explain the observed hot-streak phenomena. Instead, our preliminary analyses of scientific careers suggest that a particular combination of research strategies, namely exploration followed by exploitation, seems to be particularly powerful in predicting the beginning of a hot streak. As more data on individual careers becomes available, we may be able to identify the drivers and triggers for hot streaks, helping us answer a range of new questions. Can we anticipate the start and the end of a hot streak? Can we create an environment to facilitate and promote the onset of a hot streak, and to extend it when it emerges?

Box 1.7.1 Bursts in human activities

Short periods of elevated performance are not limited to creative achievements: bursts are also observed in a wide range of human activity patterns. For example, it is often assumed that the timing of specific human activity, like making phone calls, is random. If it is truly random, the inter-event times between consecutive events should follow an exponential distribution. The measurements indicate otherwise: The inter-event times in most human activities is well approximated by a power law distribution [132-134]. This means that the sequence of events is characterized by bursts of activities, occurring within a relatively short time frame, and occasionally there are long time gaps between two events. The “bursty” pattern of human behavior has been documented in a number of activities, from email communications to call patterns and sexual contacts. Although the burstiness and hot streaks are measured differently, these examples suggest that hot streaks have commonalities with the bursts observed in human activities, and the hot-streak phenomena may have relevance beyond human creativity.

1.7.3 What do hot streaks mean for us?

In science, future impact is critical for hiring, career advancement, research support, and other decisions. Yet, the hot streak phenomenon shows that performance can be rather uneven: scientists have substantially higher performance during a hot streak. Indeed, the timing and magnitude of hot streaks dominate a researcher’s career impact, measured by the total citations collected by all of her papers. As such, ignoring the existence of hot streaks—and their singular nature—may lead us to systematically over- or under-estimate the future impact of a career.

For example, if a hot streak comes early in a career, it would lead to a period of high impact that peaks early. That impact may diminish, however, unless a second hot streak emerges. On the other hand, if an individual has not yet experienced her hot streak, judging her career based on her current impact may underestimate her future potential. The hot streak phenomenon is also quite relevant to funding agencies, given that hot streaks and research grant both last about four years, raising the question of how can funding best maximize its impact on individual careers.

But, admittedly, implementing changes in light of the hot streak phenomenon is a challenge. It would be absurd for a junior scientist to justify his middling dossier to his tenure committee by saying, “My hot streak is coming!” Similar puzzles confront prize committees when the prize has an age threshold. The Fields Medal, the highest reward in mathematics, only recognizes mathematicians under 40. The National

Science Fund for Distinguished Young Scholars, an important milestone of a successful scientific career in China, excludes applicants over 45. Given these arbitrary age thresholds and the randomness of hot streaks, it means that a substantial portion of well-deserving candidates will miss out if they experience their big break late.

Further, our preliminary analysis also suggests that incorporating hot streaks in funding decision may not be as intuitive as it appears either. For example, we examined the relationship between the timing of NIH awards and the beginning of hot streaks for their PIs, finding that PIs are more likely to have initiated a hot streak than expected *prior to* being awarded their first R01 grant. In other words, a scientist's hot streak did not follow funding. Rather, it was funding that follows hot streaks. These findings are consistent with NIH's emphases on preliminary results, and they also appear reasonable from a funder's perspective, as people who have had a hot streak tend to be more productive (hence worthy of funding). Yet at the same time, these findings raise the question of whether—by associating funding decisions with past success—funders may miss the critical period when PIs are most creative, especially considering the fact that hot streaks are usually unique in an individual career.

Taken together, if our goal is to identify and nurture individuals who are more likely to have a lasting impact in their field, we must consider incorporating the notion of hot streaks into the calculus. If we don't, we may miss out on vitally important contributions. Yale University learned this the hard way.

We already encountered John Fenn, and his late discovery. It was at age 67, when he published a study that identified a novel electrospray ion source. It was a real breakthrough… or at least he thought so. But Yale still pushed him out the door. After he reluctantly relocated to Virginia Commonwealth University, which provided him with the lab he needed to continue his investigations, he embarked on a classical hot streak. Between 1984 and 1989, he published one study after another, ultimately developing electrospray ionization for mass spectrometry, which enabled faster, more accurate measurement of large molecules and proteins, spurring multiple innovations in cancer diagnosis and therapy. That five-year hot streak, which took place during Fenn's forced retirement, ultimately defined his career, and earned him the 2002 Nobel Prize in Chemistry. His hot streak is so evident that, you can see directly by inspecting Fig. 1.5.4. When is Fenn's most-cited paper? And when is his second highest? What about the third? They all fall within the same five-year window.

In Fenn's career, we see the most important—and uplifting—implication of the hot streak phenomenon for the many individual scientists out there striving to make their mark on the world. Remember that the conventional view, which we discussed in depth in Chapter 1.3, is this: Our best work will likely happen in our 30s or 40s, when we have a solid base of experience and the energy and enthusiasm to sustain high productivity; once we pass the mid-career point, our hopes of a breakthrough start to dim. The hot streak phenomenon, coupled with the random impact rule contradicts this, indicating, instead, that a hot streak can emerge at any stage of your career, resulting in a cluster of high impact works.

Therefore, this chapter offers hope: Each new gray hair, literal or figurative, does not by itself make us obsolete. As long as we keep putting work out into the world like Fenn—drawing cards from the deck in search of a lucky ace—our hot streak could be just around the corner.

This means that while we may not all have Einstein or Fenn's impact, if we keep trying, our own version of a miracle year may still be ahead, just out of sight.

Bibliography

1. Dennis, W., *Bibliographies of eminent scientists.* Scientific Monthly, New York, 1954.
2. Simonton, D.K., *Creative productivity: A predictive and explanatory model of career trajectories and landmarks.* Psychological Review, 1997. **104**(1): p. 66.
3. Brodetsky, S., *Newton: Scientist and man.* Nature, 1942. **150**: p. 698-699.
4. Dong, Y., et al. *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*. 2017. ACM.
5. Sinatra, R., et al., *A century of physics.* Nature Physics, 2015. **11**(10): p. 791-796.
6. Goldberger, M.L., B.A. Maher, and P.E.E. Flattau, *doctorate programs in the United States: Continuity and change*. 1995: National Academy Press.
7. Baird, L., *Departmental publication productivity and reputational quality: Disciplinary differences.* Tertiary Education and Management, 2009. **15**(4): p. 355-369.
8. Ioannidis, J.P., *Why most published research findings are false.* PLoS medicine, 2005. **2**(8): p. e124.
9. Shockley, W., *On the statistics of individual variations of productivity in research laboratories*. Proceedings of the IRE, 1957. **45**(3): p. 279-290.
10. Fronczak, P., A. Fronczak, and J.A. Hołyst, *Analysis of scientific productivity using maximum entropy principle and fluctuation-dissipation theorem.* Physical Review E, 2007. **75**(2): p. 026103.
11. Lotka, A.J., *The frequency distribution of scientific productivity.* Journal of Washington Academy Sciences, 1926.
12. de Solla Price, D., *Little science, big science*. 1963: Columbia University Press. 119.
13. Lehman, H.C., *Men's creative production rate at different ages and in different countries.* The Scientific Monthly, 1954. **78**: p. 321-326.
14. Allison, P.D. and J.A. Stewart, *Productivity differences among scientists: Evidence for accumulative advantage.* American sociological review, 1974: p. 596-606.
15. Radicchi, F. and C. Castellano, *Analysis of bibliometric indicators for individual scholars in a large data set.* Scientometrics, 2013. **97**(3): p. 627-637.
16. Barabási, A.-L., *The Formula: The Universal Laws of Success*. 2018: Hachette UK.
17. Bertsimas, D., et al., *Or forum—tenure analytics: Models for predicting research impact.* Operations research, 2015. **63**(6): p. 1246-1261.
18. Stephan, P.E., *How economics shapes science*. Vol. 1. 2012: Harvard University Press Cambridge, MA.
19. Clauset, A., S. Arbesman, and D.B. Larremore, *Systematic inequality and hierarchy in faculty hiring networks.* Science advances, 2015. **1**(1): p. e1400005.
20. Broad, W.J., *The publishing game: getting more for less.* Science, 1981. **211**(4487): p. 1137-1139.
21. Smalheiser, N.R. and V.I. Torvik, *Author name disambiguation.* Annual review of information science and technology, 2009. **43**(1): p. 1-43.
22. Ferreira, A.A., M.A. Gonçalves, and A.H. Laender, *A brief survey of automatic methods for author name disambiguation.* Acm Sigmod Record, 2012. **41**(2): p. 15-26.
23. Torvik, V.I., et al., *A probabilistic similarity metric for Medline records: A model for author name disambiguation.* Journal of the American Society for information science and technology, 2005. **56**(2): p. 140-158.
24. Hey, A.J. and P. Walters, *Einstein's mirror*. 1997: Cambridge University Press.
25. Mermin, N.D.A.V.I.D. *My Life with Landau*. 1990.
26. Hirsch, J.E., *An index to quantify an individual's scientific research output.* Proceedings of the National academy of Sciences of the United States of America, 2005. **102**(46): p. 16569-16572.
27. Van Noorden, R., *Metrics: A profusion of measures.* Nature, 2010. **465**(7300): p. 864-866.
28. Van Raan, A.F., *Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups.* scientometrics, 2006. **67**(3): p. 491-502.
29. Zhivotovsky, L. and K. Krutovsky, *Self-citation can inflate h-index.* Scientometrics, 2008. **77**(2): p. 373-375.
30. Purvis, A. *The h index: playing the numbers game*. 2006.

31. Hirsch, J.E., *Does the h index have predictive power?* Proceedings of the National Academy of Sciences, 2007. **104**(49): p. 19193-19198.
32. Cattell, J.M., *American men of science: a biographical directory*. 1910: The Science Press.
33. Lane, J., *Let's make science metrics more scientific.* Nature, 2010. **464**(7288): p. 488-489.
34. Alonso, S., et al., *hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices.* Scientometrics, 2009. **82**(2): p. 391-400.
35. Alonso, S., et al., *h-Index: A review focused in its variants, computation and standardization for different scientific fields.* Journal of Informetrics, 2009. **3**(4): p. 273-289.
36. Burrell, Q.L., *On the h-index, the size of the Hirsch core and Jin's A-index.* Journal of Informetrics, 2007. **1**(2): p. 170-177.
37. Cabrerizo, F.J., et al., *q2-Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core.* Journal of Informetrics, 2010. **4**(1): p. 23-28.
38. Jin, B., et al., *The R-and AR-indices: Complementing the h-index.* Chinese science bulletin, 2007. **52**(6): p. 855-863.
39. Kosmulski, M., *A new Hirsch-type index saves time and works equally well as the original h-index.* ISSI newsletter, 2006. **2**(3): p. 4-6.
40. Egghe, L., *An improvement of the h-index: the g-index.* ISSI newsletter, 2006. **2**(1): p. 8-9.
41. Egghe, L., *Theory and practise of the g-index.* Scientometrics, 2006. **69**(1): p. 131-152.
42. Dorogovtsev, S.N. and J.F.F. Mendes, *Ranking scientists.* Nature Physics, 2015. **11**(11): p. 882-883.
43. Radicchi, F., S. Fortunato, and C. Castellano, *Universality of citation distributions: Toward an objective measure of scientific impact.* Proceedings of the National Academy of Sciences, 2008. **105**(45): p. 17268-17272.
44. Kaur, J., F. Radicchi, and F. Menczer, *Universality of scholarly impact metrics.* Journal of Informetrics, 2013. **7**(4): p. 924-932.
45. Sidiropoulos, A., D. Katsaros, and Y. Manolopoulos, *Generalized Hirsch h-index for disclosing latent facts in citation networks.* Scientometrics, 2007. **72**(2): p. 253-280.
46. Hirsch, J., *An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship.* Scientometrics, 2010. **85**(3): p. 741-754.
47. Hirsch, J.E., *h α: An index to quantify an individual's scientific leadership.* Scientometrics, 2019. **118**(2): p. 673-686.
48. Schreiber, M., *A modification of the h-index: The h m-index accounts for multi-authored manuscripts.* Journal of Informetrics, 2008. **2**(3): p. 211-216.
49. Egghe, L., *Mathematical theory of the h-and g-index in case of fractional counting of authorship.* Journal of the American Society for Information Science and Technology, 2008. **59**(10): p. 1608-1616.
50. Galam, S., *Tailor based allocations for multiple authorship: a fractional gh-index.* Scientometrics, 2011. **89**(1): p. 365.
51. Tscharntke, T., et al., *Author sequence and credit for contributions in multiauthored publications.* PLoS Biol, 2007. **5**(1): p. e18.
52. Ausloos, M., *Assessing the true role of coauthors in the h-index measure of an author scientific impact.* Physica A: Statistical Mechanics and its Applications, 2015. **422**: p. 136-142.
53. Liu, X.Z. and H. Fang, *Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution.* Journal of Informetrics, 2012. **6**(4): p. 557-565.
54. Hu, X., R. Rousseau, and J. Chen, *In those fields where multiple authorship is the rule, the h-index should be supplemented by role-based h-indices.* Journal of Information Science, 2010. **36**(1): p. 73-85.
55. Google. Available from: https://scholar.google.com.
56. Radicchi, F., et al., *Diffusion of scientific credits and the ranking of scientists.* Physical Review E, 2009. **80**(5): p. 056103.
57. Abbott, A., et al., *Metrics: Do metrics matter?* Nature News, 2010. **465**(7300): p. 860-862.
58. Pavlou, M. and E.P. Diamandis, *The athletes of science.* Nature, 2011. **478**(7369): p. 419-419.
59. Kuhn, T.S., *The structure of scientific revolutions*. 1962: University of Chicago press.
60. Merton, R.K., *The Matthew effect in science.* Science, 1968. **159**(3810): p. 56-63.
61. Simcoe, T.S. and D.M. Waguespack, *Status, quality, and attention: What's in a (missing) name?* Management Science, 2011. **57**(2): p. 274-290.
62. Tomkins, A., M. Zhang, and W.D. Heavlin, *Reviewer bias in single-versus double-blind peer review.* Proceedings of the National Academy of Sciences, 2017. **114**(48): p. 12708-12713.

63. De Ranieri, E., et al., *Analysis of uptake and outcome in author-selected single-blind vs doubleblind peer review at nature journals. Presentation at the International Congress on Peer-review and Scientific Publication.[cited 2018 October 15]*. 2017.
64. Blank, R.M., *The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review*. The American Economic Review, 1991: p. 1041-1067.
65. Petersen, A.M., et al., *Reputation and Impact in Academic Careers.* Proceedings of the National Academy of Science USA, 2014. **111**: p. 15316-15321.
66. Cole, S., *Age and scientific performance.* American journal of sociology, 1979: p. 958-977.
67. Newman, M., *Networks: an introduction*. 2010: Oxford University Press.
68. Barabási, A.-L., *Network science*. 2015: Cambridge University.
69. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
70. Mazloumian, A., et al., *How citation boosts promote scientific paradigm shifts and nobel prizes.* PloS one, 2011. **6**(5): p. e18975.
71. Fang, F.C., R.G. Steen, and A. Casadevall, *Misconduct accounts for the majority of retracted scientific publications.* Proceedings of the National Academy of Sciences, 2012. **109**(42): p. 17028-17033.
72. Lu, S.F., et al., *The retraction penalty: evidence from the Web of Science.* Scientific reports, 2013. **3**.
73. Azoulay, P., et al., *Retractions.* Review of Economics and Statistics, 2014(0).
74. Azoulay, P., A. Bonatti, and J.L. Krieger, *The Career Effects of Scandal: Evidence from Scientific Retractions*. 2015, National Bureau of Economic Research.
75. Jin, G.Z., et al., *The reverse Matthew effect: catastrophe and consequence in scientific teams*. 2013, National Bureau of Economic Research.
76. Merton, R.K., *Singletons and multiples in scientific discovery: A chapter in the sociology of science.* Proceedings of the American Philosophical Society, 1961: p. 470-486.
77. Azoulay, P., T. Stuart, and Y. Wang, *Matthew: Effect or fable?* Management Science, 2013. **60**(1): p. 92-109.
78. Garfield, E. and A. Welljams-Dorof, *Of Nobel class: A citation perspective on high impact research authors.* Theoretical Medicine, 1992. **13**(2): p. 117-135.
79. Azoulay, P., *Research efficiency: Turn the scientific method on ourselves.* Nature, 2012. **484**(7392): p. 31-32.
80. Restivo, M. and A. Van De Rijt, *Experimental study of informal rewards in peer production.* PloS one, 2012. **7**(3): p. e34358.
81. van de Rijt, A., et al., *Field experiments of success-breeds-success dynamics.* Proceedings of the National Academy of Sciences, 2014. **111**(19): p. 6934-6939.
82. Alberts, B., et al., *Opinion: Addressing systemic problems in the biomedical research enterprise.* Proceedings of the National Academy of Sciences, 2015. **112**(7): p. 1912-1913.
83. Kaiser, J., *Biomedical research. The graying of NIH research.* Science (New York, NY), 2008. **322**(5903): p. 848-849.
84. Beard, G.M., *Legal responsibility in old age.* Russells, New York, 1874: p. 5-42.
85. Lehman, H.C., *Age and achievement.* Princeton, NJ: Princeton, 1953.
86. Dennis, W., *Age and productivity among scientists.* Science, 1956. **123**: p. 724-725.
87. Dennis, W., *Creative productivity between the ages of 20 and 80 years.* Journal of Gerontology, 1966. **21**(1): p. 1-8.
88. Jones, B.F., *Age and great invention.* The Review of Economics and Statistics, 2010. **92**(1): p. 1-14.
89. Jones, B., E.J. Reedy, and B.A. Weinberg, *Age and scientific genius*. 2014, National Bureau of Economic Research.
90. Usher, A.P., *A History of Mechanical Inventions: Revised Edition*. 1954: Courier Corporation.
91. Weitzman, M.L., *Recombinant growth.* Quarterly Journal of Economics, 1998: p. 331-360.
92. Uzzi, B., et al., *Atypical combinations and scientific impact.* Science, 2013. **342**(6157): p. 468-472.
93. Ericsson, K.A., R.T. Krampe, and C. Tesch-Römer, *The role of deliberate practice in the acquisition of expert performance.* Psychological review, 1993. **100**(3): p. 363.
94. Ericsson, K.A. and A.C. Lehmann, *Expert and exceptional performance: Evidence of maximal adaptation to task constraints.* Annual review of psychology, 1996. **47**(1): p. 273-305.
95. Ericsson, K.A., et al., *The Cambridge handbook of expertise and expert performance*. 2006: Cambridge University Press.

96. Pelz, D.C. and F.M. Andrews, *SCIENTISTS IN ORGANIZATIONS: PRODUCTIVE CLIMATES FOR RESEARCH AND DEVELOPMENT.* 1966.
97. Bayer, A.E. and J.E. Dutton, *Career age and research-professional activities of academic scientists: Tests of alternative nonlinear models and some implications for higher education faculty policies.* The Journal of Higher Education, 1977: p. 259-282.
98. Blackburn, R.T., C.E. Behymer, and D.E. Hall, *Research note: Correlates of faculty publications.* Sociology of Education, 1978: p. 132-141.
99. Matthews, K.R., et al., *The aging of biomedical research in the United States.* 2011.
100. Adams, C.H.A.R.L.E.S.W., *The age at which scientists do their best work.* Isis, 1946: p. 166-169.
101. Zuckerman, H., *Scientific elite: Nobel laureates in the United States*. 1977: Transaction Publishers.
102. Simonton, D.K., *Career landmarks in science: Individual differences and interdisciplinary contrasts.* Developmental Psychology, 1991. **27**(1): p. 119.
103. Jones, B.F. and B.A. Weinberg, *Age dynamics in scientific creativity.* Proceedings of the National Academy of Sciences, 2011. **108**(47): p. 18910-18914.
104. Jones, B.F., *The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?* The Review of Economic Studies, 2009. **76**(1): p. 283-317.
105. Jones, B.F., *As science evolves, how can science policy?*, in *Innovation Policy and the Economy, Volume 11*. 2011, University of Chicago Press. p. 103-131.
106. Machlup, F., *The production and distribution of knowledge in the United States*. Vol. 278. 1962: Princeton university press.
107. Fortunato, S., et al., *Growing time lag threatens Nobels.* Nature, 2014. **508**(7495): p. 186-186.
108. Cassidy, D.C., *Uncertainty. The life and science of Werner Heisenberg.* New York: Freeman, 1992, 1992. **1**.
109. Weinberg, B.A. and D.W. Galenson, *Creative careers: The life cycles of Nobel laureates in economics*. 2005, National Bureau of Economic Research.
110. Rappa, M. and K. Debackere, *Youth and scientific innovation: The role of young scientists in the development of a new field.* Minerva, 1993. **31**(1): p. 1-20.
111. Packalen, M. and J. Bhattacharya, *Age and the Trying Out of New Ideas*. 2015, National Bureau of Economic Research.
112. Galenson, D.W., *Painting outside the lines: Patterns of creativity in modern art*. 2009: Harvard University Press.
113. Galenson, D.W., *Old masters and young geniuses: The two life cycles of artistic creativity*. 2011: Princeton University Press.
114. Hull, D.L., P.D. Tessner, and A.M. Diamond, *Planck's principle.* Science, 1978. **202**(4369): p. 717-723.
115. Azoulay, P., J.S. Zivin, and J. Wang, *Superstar Extinction.* Quarterly Journal of Economics, 2010. **125**(2): p. 549-589.
116. Sinatra, R., et al., *Quantifying the evolution of individual scientific impact.* Science, 2016. **354**(6312): p. aaf5239.
117. Liu, L., et al., *Hot streaks in artistic, cultural, and scientific careers.* Nature, 2018: p. 1.
118. Simonton, D.K., *Creative productivity, age, and stress: a biographical time-series analysis of 10 classical composers.* Journal of Personality and Social Psychology, 1977. **35**(11): p. 791.
119. Simonton, D.K., *Quality, quantity, and age: The careers of ten distinguished psychologists.* The International Journal of Aging & Human Development, 1985.
120. Simonton, D.K., *Genius, creativity, and leadership: Historiometric inquiries*. 1984: Harvard Univ Pr.
121. Simonton, D.K., *Scientific genius: A psychology of science*. 1988: Cambridge Univ Pr.
122. Li, J., et al., *Nobel laureates are almost the same as us.* Nature Reviews Physics, 2019. **1**(5): p. 301.
123. Jones, B.F., N.J.D. Kim, and M.J. Miranda, *Age and High-Growth Entrepreneurship.* 2018.
124. Azoulay, P., et al., *Research: the average age of a successful startup founder is 45*. 2019.
125. Powdthavee, N., Y.E. Riyanto, and J.L. Knetsch, *Lower-rated publications do lower academics' judgments of publication lists: Evidence from a survey experiment of economists.* Journal of Economic Psychology, 2018. **66**: p. 33-44.
126. Gilovich, T., R. Vallone, and A. Tversky, *The Hot Hand in Basketball - on the Misperception of Random Sequences.* Cognitive Psychology, 1985. **17**(3): p. 295-314.
127. Miller, J. and A. Sanjurjo, *Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers*. 2015, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.
128. Ayton, P. and I. Fischer, *The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?* Memory & Cognition, 2004. **32**(8): p. 1369-1378.

129. Rabin, M. and D. Vayanos, *The Gambler's and Hot-Hand Fallacies: Theory and Applications.* Review of Economic Studies, 2010. **77**(2): p. 730-778.
130. Xu, J.M. and N. Harvey, *Carry on winning: The gamblers' fallacy creates hot hand effects in online gambling.* Cognition, 2014. **131**(2): p. 173-180.
131. Csapo, P. and M. Raab, *"Hand down, Man down." Analysis of Defensive Adjustments in Response to the Hot Hand in Basketball Using Novel Defense Metrics (vol 9, e114184, 2014).* Plos One, 2015. **10**(4).
132. Barabasi, A.-L., *The origin of bursts and heavy tails in human dynamics.* Nature, 2005. **435**(7039): p. 207.
133. Vázquez, A., et al., *Modeling bursts and heavy tails in human dynamics.* Physical Review E, 2006. **73**(3): p. 036127.
134. Barabási, A.-L., *Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades*. 2010: Penguin.