# Long live the scientists: Tracking the scientific fame of great minds in physics

Guoyan Wang [a], Guangyuan Hu [b], Chuanfeng Li [c], Li Tang [d],*

[a] *Department of Science and Technology Communication and Policy, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui Province, 230026, China*
[b] *School of Public Economics and Public Administration, Shanghai University of Finance Economics, 777 Guoding Road, Shanghai, 200433, China*
[c] *Key Lab of Quantum Information, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui Province, 230026, China*
[d] *School of International Relations and Public Affairs, Fudan University, 220 Handan Road, Shanghai, 200433, China*

A B S T R A C T

This study utilizes global digitalized books and articles to examine the scientific fame of the most influential physicists. Our research reveals that the greatest minds are gone but not forgotten. Their scientific impacts on human history have persisted for centuries. We also find evidence in support of own-group fame preference, i.e., that the scientists have greater reputations in their home countries or among scholars sharing the same languages. We argue that, when applied appropriately, Google Books and Ngram Viewer can serve as promising tools for altmetrics, providing a more comprehensive picture of the impacts scholars and their achievements have made beyond academia.

© 2018 Elsevier Ltd. All rights reserved.

*Some say that a man dies three times. The first time is when his heart stops beating and he dies physically. The second is when people come to his funeral and his identity is erased from society. The third time is when nobody on the earth remembers him anymore. Then he is really dead.*

"Dragon Raja" by Lee Yeongdo

## 1. Introduction

Books are the stepping stones to human progress. According to UNESCO (United Nations Educational, Scientific and Cultural Organization), the number of estimated published books in 2017 alone is up to 2.2 million.[1] Such a large collection is undoubtedly a rich archive of human history and civilization. Yet, as one of the most telling embodiments of knowledge stock and advancement, books have not captured sufficient attention in quantitative research evaluation.

Fortunately, with access to Google Books and the Google Books tool Ngram Viewer, scholars are now able to trace cultural evolution on a long time scale based on digitalized texts and trillions of words. This application of high-throughput data collection to study human culture can be traced back to Michel et al. (2011). In this pioneering study, the authors utilized the Google Books corpus and conducted text-based statistical analysis to trace cultural trends. That innovative research method soon captured academia's attention and was adopted in the arenas of digital history (Sternfeld, 2011), the history of science

---

* Corresponding author.
  *E-mail address:* litang@fudan.edu.cn (L. Tang).
[1] Data source: http://www.worldometers.info/books/. Accessed on January 18, 2018.

(Laubichler, Maienschein, & Renn, 2013), economics (Roth, 2013), social psychology (Acerbi, Lampos, Garnett, & Bentley, 2013; Greenfield, 2013; Zeng & Greenfield, 2015), and cultural psychology (Pettit, 2016).

Google Books has also been utilized to assess the fame of great scientists throughout history. Based on the word frequency of people's full names mentioned in books, Bohannon's Science Hall of Fame was built a as an objective evaluation of scientific fame over centuries (Bohannon, 2011a, 2011b). Moving beyond previous work, this paper utilizes both Google Books, which covers 36 million global digital books, and Google Scholar, which indexes 91 million academic items,[2] to examine the scientific fame of top physicists. We particularly focus on and compare two of the greatest physicists, Isaac Newton and Albert Einstein, depicting their fame evolution over centuries and exploring what they are famous for.

Our research reveals that the great minds are gone but not forgotten. Early scientists are still on the public's lips in modern society. Their scientific impacts on human history have persisted for centuries. This holds true for other prominent physicists as well. We also found that while Einstein's scientific fame has exceeded that of Newton among intellectuals since the mid-20th century worldwide, there is a different pattern of fame when the own-group preference is differentiated by the language of digitalized corpus. The computational analysis confirms that the influence of Einstein is largely related to his two contributions on general relativity and quantum theory, while the most frequently mentioned scientific achievements of Newton are the law of universal gravitation and the laws of motion.

This paper makes the following contributions to the literature. To begin with, this is the first attempt, within our best knowledge, to explore scientific fame based on the combination of both books and articles indexed by Google. In addition to depicting the evolution of fame of great minds, we also explore their most accredited achievements based on co-occurence analysis. Second, our study contributes to the discussion on the expected role of scientists in science promotion. By comparing the indicator for scientific impact (i.e., work being cited by scholarly publications) and the indicator for scientific fame (i.e., name being mentioned in books), our study sheds some light on how to gauge scientists' contributions beyond academia. We argue that, when applied appropriately, Google Books and Google Ngram Viewer can serve as promising tools for altmetrics, providing a more comprehensive picture of the impacts scholars and their achievements have made on society.

The rest of the paper is structured as follows. In the next section, we delineate our method and case selection justifications. Section 3 presents our analysis. In Section 4, following a summary of main findings, we conclude our paper discussing limitations and future research venues.

## 2. Method and data

### 2.1. Notion and measurement

There is no agreed-upon definition of scientific fame or its measurement. The term can be traced back to the book *The Life of Sir Charles Linnæus*, a biography of a Swedish botanist, physician, and zoologist whose fame is centered on his enduring achievement of binomial nomenclature (Stöver, Stöver, Trapp, Stoever, & Linne´, 2007). Yet many scientists are ordinary folks and are little known to the public (Astin, 1957; Menard, 1971; Merton, 1970). Some scholars have argued that the fame of scientists should be confined to professional achievements, while others believe scientific fame goes beyond academia (Bohannon, 2011a; Menard, 1971). Feist (2016) noted that regardless of either intrinsic or extrinsic research, the assessment of fame in science ultimately rests on productivity and its impact on advancing the research front. Previous studies often used being elected to prestigious societies or winning research awards or prizes as proxy indicators of scientific recognition or reputation (Bronk, 1976; Youtie, Rogers, Heinze, Shapira, & Tang, 2013 etc.). Instead of relying on the subjective judgment of panel experts, Bohannon's Science Hall of Fame (2011a) innovatively uses the appearance of people's names in books to capture scientists' influence across different domains throughout history. This is the approach we adopt in tracking and recording the fame of great scientific minds and their achievements.

### 2.2. Case selection

The focus domain in this work is physics. Among the myriad scientists, we purposely choose Isaac Newton and Albert Einstein for illustration based on the following considerations.

To begin with, both Newton and Einstein are in the field of physics, which makes their comparison relatively free from discipline differences of publication distribution. And given that both names consist of two words with similar length, the quality of their retrievals is fairly comparative. As the two most influential physicists in the history of science (Baker, 1984; Whittaker et al., 1943), Newton and Einstein are appropriate candidates for evaluating the historical fame of individuals. Finally, who is more influential has been a topic attracting much attention in the global scientific community. The debate has not been settled for more than half a century (Gribbin, 1987; Graneau & Graneau, 1993). Successor scientists have commemorated them on special anniversaries, such as "Science 1943: Aristotle, Newton, Einstein, the three-hundredth anniversary of Newton's birth" and "New Scientist 1987: Newton vs. Einstein, the three-hundredth anniversary of Newton's theory of gravity." In 2005, the Year of World Physics and also the Centenary of Einstein's Special Relativity Theory, the UK

---

[2] The types of academic items Google Scholar includes are research articles, books, patents, case laws, and citations. In this research our search excludes patents and citations.

**Table 1**
Coverage of Google Books and Google Scholar.

| Google corpus | Total returned hits | After 2000 | 20th century | 19th century | 18th century | 17th century | Before 1600 |
|---|---|---|---|---|---|---|---|
| Books | 36,540,000 | 14,600,000 | 14,700,000 | 5,680,000 | 808,000 | 319,000 | 433,000 |
| Scholar | 91,258,000 | 43,061,000 | 46,935,600 | 1,180,630 | 80,300 | 2 | 0 |

*Note*: This search was conducted through the library at London University in July 2015. For Google Books, the command we used is "allinurl:books" at https://books.google.com. For Google Scholar we set up the custom range of years in the advanced search.[a]

[a]It needs to be pointed out that depending on the search date and location, the returned hits vary. Take Google Books, for example. We repeated the same search in July 2018, and the returned hits are not 36,540,000 as in July 2015 in the UK but are 38,400,000 in China and 39,300,000 in South Korea. This pattern also holds for Google Scholar. As shown in Table 1, the coverage of Google Scholar was 43,061,000 articles published over the period of 2000–2015 when the search was conducted at London University in July 2015. In August 2018, its coverage of the same period extends to 46,840,000 and 48,720,000 articles when searched in China and South Korea, respectively. Our experiences echo previous findings on the coverage scope, growth rate, and indexing quality of Google Books and Google Scholar (Abdullah and Thelwall, 2014; Fagan, 2017; Halevi et al., 2017). For an extensive review of the pros and cons of Google Books and Google Scholar as well as their utilities in research assessment, please refer to Google Scholar Digest at http://googlescholardigest.blogspot.com/p/bibliography.html.
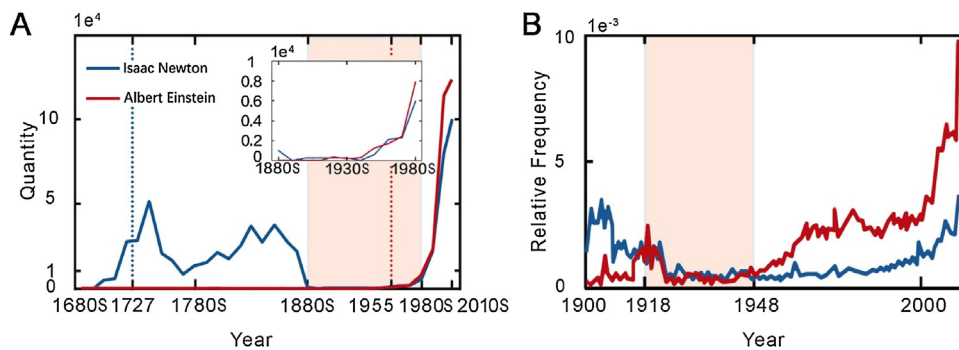


**Fig. 1.** Global fame of Newton and Einstein. Accessed in January 2016. Forty-nine languages with the same name spelling as in English are included. Data source: Panel A is Google Books search. The time coverage of the main graph is from 1680s to 2010s with an inset zooming in on the period of 1880s–1980s; Panel B is Google Scholar search. The time coverage is from 1900 to 2012.

Royal Society conducted polls of both academia and the public. The results showed that both Royal Society members and the British netizens surveyed considered Isaac Newton, a British scientist, to have greater influence on both science and humankind than Albert Einstein (*The Royal Society News*, 2005). It would be interesting to find out whether that also holds true beyond the geography of the UK.

### 2.3. Data

Our main datasets are Google digital books and scholarly articles. As shown in Table 1, the Google corpus used in this study consists of over 36 million global digital books and 91 million articles published since the 16th century.

To gain a complete picture of these physicists' fame worldwide, we took a variety of languages into consideration. We first tested our search terms in 57 different languages using the Google Books search engine based on trials and errors.[3] The volumes and shares of the written languages of these books are displayed in Table 2. As shown, the dataset we analyzed covers more than 126 thousand books mentioning Newton and 149 thousand books mentioning Einstein since the 20th century. This finding seems opposite to that of the UK polls in 2005. Table 2 also reveals that, similar to other publication datasets, the coverage of Google Books is also biased in language (Halevi et al., 2017; Abdullah and Thelwall, 2014; Liu et al., 2015, 2018). Books written in Chinese, Russian, Korean, and other non-English languages are highly underrepresented in the Google Books corpus. This caveat needs to be borne in mind when using this search engine for analysis.

### 3. Analysis

#### 3.1. Global fame over time

Fig. 1 depicts the quantity of digitalized books and academic items mentioning Newton's and Einstein's full names. It needs to be pointed out that our returned hits consist of two scenarios: books mentioning Newton or Einstein and books written by either of them. But given the number of returned hits and the number of books even the most prolific scholar

---

[3] The variations of name order (e.g., "Isaac Newton" and "Newton, Isaac") were taken into consideration.

**Table 2**
Retrieved hits in Google Books by different languages.

| Scientist | Search word | Retrieved hits in 57 languages | 2000–2015 (prop.) | 1900–1999 (prop.) |
|---|---|---|---|---|
| **Isaac Newton** (1643–1727) | Isaac Newton | Records in 49 languages with the same name spelling as in English | 95.78% | 98.18% |
| | Исаак Ньютон | Russian | 2.01% | 1.28% |
| | 아이작 뉴턴 | Korean | 1.17% | 0.12% |
| | アイザック・ニュートン | Japanese | 0.51% | 0.31% |
| | "艾萨克·牛顿"OR"艾薩克·牛頓" | Chinese (simplified or traditional) | 0.21% | 0.06% |
| | إسحاق نيوتن | Arabic | 0.18% | 0.04% |
| | אייזק ניוטון | Hebrew | 0.15% | 0.02% |
| | আইজাক নিউটন | Bengali | 0.01% | 0.00% |
| | आइज़ैक न्यूटन | Hindi | 0.00% | 0.00% |
| | **Total retrieved hits of Newton** | | **125593** | **36434** |
| **Albert Einstein** (1879–1955) | Albert Einstein | Records in 49 languages with the same name spelling as in English | 96.25% | 96.78% |
| | Альберт Эйнштейн | Russian | 2.35% | 2.08% |
| | אלברט איינשטיין | Hebrew | 0.42% | 0.62% |
| | アルバート・アインシュタイン | Japanese | 0.35% | 0.22% |
| | 알버트 아인슈타인 | Korean | 0.26% | 0.08% |
| | "阿尔伯特·爱因斯坦"OR"阿爾伯特·愛因斯坦" | Chinese (simplified or traditional) | 0.21% | 0.17% |
| | البرت اينشتاين | Arabic | 0.05% | 0.06% |
| | अल्बर्ट आइंस्टीन | Hindi | 0.01% | 0.00% |
| | আলবার্ট আইনস্টাইন | Bengali | 0.00% | 0.00% |
| | **Total retrieved hits of Einstein** | | **149169** | **38154** |

Data source: Google Books search. Data accessed in January 2016.
*Note*: In the other 49 languages such as German, French, Italian, Spanish, and Portuguese, both name spellings are exactly the same as in English.

can write, it is reasonable to believe that the majority of the retrieved books are those mentioning them. Section 4 discusses this in more detail. Blue denotes Newton while red represents Einstein. The dotted lines mark the deceased years of the two great scientists, respectively. Please note that any key word search is only effective for accessible content. In our case, only Google Books with a preview available can be searched for names in the text.

In Fig. 1A, the number in the upper left ($1e^4=10^4$) is the unit of the absolute number of books indexed in the Google Books search. For Newton, the most glorious period of his historical influence is between 1680 and 1880, when records mentioning Newton amount to more than 60% of all records mentioning him. As demonstrated, the deaths of these great scientists do not dim their glory. On the contrary, the popularity of both physicists has risen dramatically since the 1980s. One may wonder what happened in the period of 1880–1980, when relatively fewer books talked about Newton and Einstein. One speculation is that two dark clouds appeared in the world of physics at the end of the 19th century: the Michelson–Morley experiment and the ultraviolet catastrophe in Planck's law (Kelvin, 1901). They overturned the framework of the old theory system, leading to the birth of quantum and relativity theories (Hoover, 1977; Sanghera, 2011). This, when combined with the breakout and aftermath of World Wars I and II, partially explains the silence of Newton's voice from 1880 to 1980.

Fig. 1B maps the scientists' influences in academia via a Google Scholar search engine. The y values, i.e. relative counts, were computed by dividing the number of Google Scholar indexed articles mentioning the scientist in a given year by the total number of Google Scholar indexed articles in that year. As shown, Newton leads in fame until 1918, then both scientists parallel for 30 years. The year of 1948 is the watershed. Since then, Einstein has gained more popularity than Newton worldwide.

One key message conveyed by Fig. 1 is that globally Einstein seems to enjoy greater fame than Newton, which is contrary to the results of the 2005 UK polls mentioned previously. If the findings of both the surveys and the computational analysis are valid, one possible explanation for this discrepancy is that scientific fame varies by location. People are more recognized
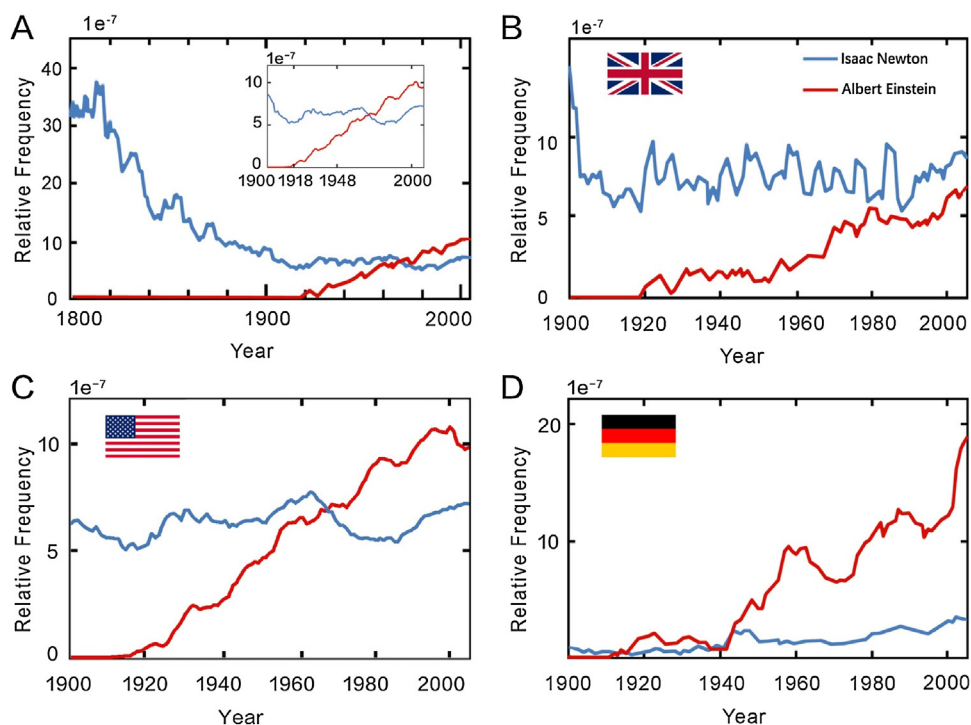
**Fig. 2.** Scientific fame by language.
Panel A: Only books in English (both British English and American English) are considered.
Panel B: Only books in British English are considered.
Panel C: Only books in American English are considered.
Panel D: Only books in German are considered.
The time coverage is from 1800 to 2008 for Panel A and from 1900 to 2008 for Panels B–D.
Data source: Google Books Ngram Viewer. Accessed in January 2016.

within their own group (Egghe & Rousseau, 2004; Tang, Shapira, & Youtie, 2015), be it by country or by shared language. To test this speculation, we utilized the language option of Google Books Ngram Viewer and examined the physicists' historical influence by various languages.[4]

### 3.2. Global fame moderated by language

Fig. 2 was generated with Google Books Ngram Viewer with a default smoothing of three.[5] The four panels illustrate the scientific fame of Newton and Einstein in books written in British and American English together, British English alone, American English alone, and German, respectively. Both scientists seem to be preferred in their own groups. In the British English case, Newton has consistently enjoyed more popularity than Einstein throughout the last 100 years (Panel B). This finding is consistent with 2005 UK polls results. In sharp contrast, Einstein, a German-origin physicist who later immigrated to the US, has been mentioned far more than Newton since the mid or late 20th century in British and American English together, American English alone, and German books (Fig. 2A, C, and D). This comparison of citing book language uncovers a high relevance of own-group preference, or approximately location of fame.

---

[4] Since 2010 the Google Books Ngram Viewer has been available at http://books.google.com/ngrams. Its source, i.e. Google Books, was generated through Partner Program and the Library Project (please refer to https://www.google.com/intl/en/googlebooks/about/). These millions of books published between 1500 and 2008 were digitally scanned and the corpus was winnowed. For more details please see Michel et al. (2011) and Twenge, Campbell, and Gentile, (2012). The searchable corpora with the Google Books Ngram Viewer were developed by the Google Team and can be downloaded at http://storage.googleapis.com/books/ngrams/books/datasetsv2.html. Please note that Google Books Ngram Viewer cannot be used to determine how often a search term occurs in a book.

[5] The default smoothing parameter of Google Books Ngram Viewer is set at 3. Taking the year of 2000 as an example, a smoothing of 3 means that the data shown for 2000 are a mathematical average of original data for 2000 plus three values before and three after 2000 (i.e., a moving average of seven-year data, including 1997, 1998, 1999, 2000, 2001, 2002, and 2003). This setting avoids spikes and makes trends more apparent (Michel et al., 2011; Greenfield, 2013). More information on graphs generated by Google Books Ngram Viewer is available at https://books.google.com/ngrams/info.
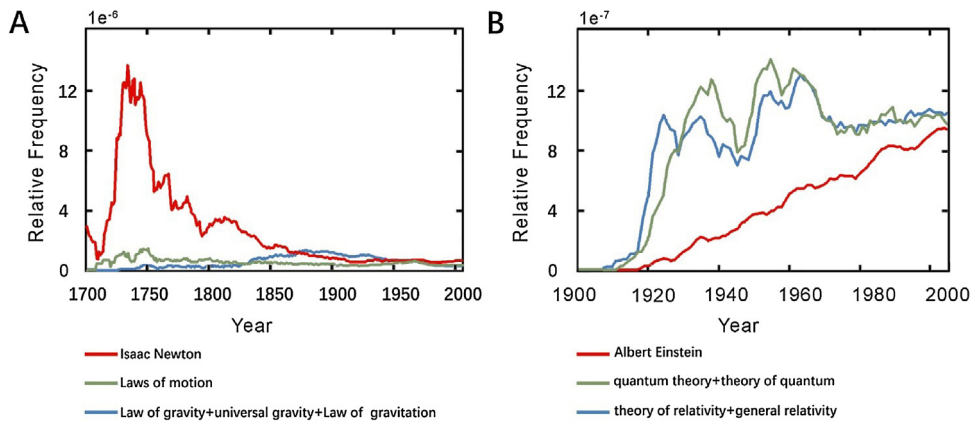
**Fig. 3.** The major scientific contributions of Newton and Einstein.
The time coverage is from 1700 to 2008 for Panel A and from 1900 to 2008 for Panel B.
Data source: Google Books Ngram Viewer. Accessed in December 2017.

**Table 3**
Co-occurrence analysis of contributions and physicists.

| Key words | 1900–2015 | |
| --- | --- | --- |
| | Records in Books | % of Total |
| Isaac Newton | 156,100 | / |
| Newton and gravity(including law of gravity, universal gravity, universal gravitation, law of gravitation) | 23,700 | 15.2% |
| Newton and calculus | 12,880 | 7.9% |
| Newton and laws of motion | 10,440 | 6.7% |
| Albert Einstein | 178,200 | / |
| Einstein and relativity (including general relativity and special relativity) | 50,130 | 28.1% |
| Einstein and quantum | 30,020 | 16.9% |
| Einstein and photoelectric | 4,200 | 2.4% |
| Einstein and mass-energy | 2,420 | 1.4% |

Data source: Google Books search. Data accessed in January 2016.

### 3.3. Achievements they are famous for

Scientists are commemorated for various reasons. Hawking and Israel (1989) note that the greatest contributions made by Newton in the history of science are the law of universal gravitation and the three laws of motion. In addition to physics, Newton also contributed to the fields of mathematics, astronomy, and the philosophy of nature (Newton, Cohen, & Schofield, 1959; Newton, 1999). In comparison, the most notable achievements of Einstein are quantum mechanics and general relativity (Torre, Daleo, & García-Mata, 2000; Whittaker et al., 1943). Then, a question arises: what are they famous for as evidenced by the Google corpus?

As demonstrated in Fig. 3A, the contributions of Newton surpass the Laws of Motion and the Law of Gravity. This is particularly true prior to mid-1800. As shown the name of Isaac Newton is mentioned in Google Books far more than his most accredited physics theories are.

One may notice the close intertwining or convergence of lines representing scientific achievements and the scientists' names in later periods. But the approximately same frequency of mentioning the achievements and the scientists' names does not mean that the mentions appear simultaneously within the same corpus books. So, we next conducted co-occurrence analysis to examine to what extent these scientists' most accredited achievements contribute to their enduring reputations (Table 3).

As depicted in Table 3, books mentioning Isaac Newton often deal with the law of universal gravity (15%) and calculus (8%). The influence of Albert Einstein is highly correlated to the theory of relativity (about 28%) and quantum theory (almost 17%). This, on one hand, reveals what achievements contribute to the glory of Newton and Einstein; on the other hand, the residue (i.e., the uncorrelated proportion) suggests other elements (such as other achievements or even anecdotes) exist.[6]

---

[6] Our manual examination of the remaining books mentioning Newton but not gravity, gravitation, calculus, or motion suggests that many are related to natural philosophy, alchemy, and theology. In addition to mentions of relativity, quantum, photoelectric and mass-energy, books mentioning Einstein also discuss his ethical and philosophical views on cosmic religion, his educational philosophy, and even his private life.

**Table 4**
Top 20 physicists ranked by scientific fame.

| Rank | Name of Physicist | Nationality of Origin | Year of Birth | Rec. of being mentioned in Books |
|---|---|---|---|---|
| 1 | Albert Einstein | German | 1879 | 145,000 |
| 2 | Max Planck | German | 1858 | 129,000 |
| 3 | Isaac Newton | UK | 1643 | 121,000 |
| 4 | Blaise Pascal | France | 1623 | 118,000 |
| 5 | Galileo Galilei | Italy | 1564 | 96,100 |
| 6 | Stephen Hawking | UK | 1942 | 82,800 |
| 7 | Niels Bohr | Denmark | 1885 | 80,100 |
| 8 | Joseph Henry | USA | 1797 | 79,100 |
| 9 | Richard Feynman | German | 1918 | 65,300 |
| 10 | Werner Heisenberg | German | 1901 | 55,400 |
| 11 | Michael Faraday | UK | 1791 | 49,500 |
| 12 | James Watt | UK | 1736 | 47,900 |
| 13 | Clerk Maxwell | UK | 1831 | 43,100 |
| 14 | Enrico Fermi | Italy | 1901 | 41,400 |
| 15 | William Thomson | UK | 1824 | 40,700 |
| 16 | Robert Hooke | UK | 1635 | 40,200 |
| 17 | Max Born | German | 1882 | 38,000 |
| 18 | Wolfgang Pauli | Austria | 1900 | 37,200 |
| 19 | Robert Oppenheimer | USA | 1904 | 36,400 |
| 20 | Thomas Young | UK | 1773 | 36,400 |

Data source: Google Books search. Data accessed in January 2016. Time coverage: 2000–2015.

### 3.4. Other great minds in physics

We extended our analysis to other great physicists in history. We first compiled a list of 234 of the most influential scientists based on the following four books:

> *The 100 Most Influential Scientists of All Time* (Rogers, 2009)
> *The 100 Most Influential Scientists* (Simmons, 2009)
> *Scientists Who Changed the World* (Tiner, 2002)
> *The Most Influential Scientists in Human Civilization* (Luo, 2012)

After removing scientists born before the year 1 BCE, such as Aristotle and Plato, we next searched each name in Wikipedia and allocated the scientists to different research fields. We then retrieved their name appearance frequency in all Google Books published after 2000. The top 20 physicists listed in Table 4 show that in the field of physics, Albert Einstein, Max Planck, Isaac Newton, Blaise Pascal, and Galileo Galilei are the top five most influential and best-known physicists in the 21st century.

Table 4 lends further support that great scientists are gone but not forgotten.[7] As shown, the early scientists such as Newton, Pascal, and Galilei are still on the public's lips in contemporary society. Undoubtedly, the achievements of these scientific pioneers have become the cornerstones of human scientific knowledge. But it is also true that besides scientists' contributions to scientific advancement, their scientific spirit, anecdotes, and other special attributes may remain as well. Students and the public learn about the great scientists via scholarly publications, textbooks, and popular science books. When a scientist becomes a public figure, his life experiences and achievements are often adapted into various biographies and even enlightening stories. In stories such as "Newton and His Apple," the scientific spirit has been used to encourage and to inspire each succeeding generation for hundreds, if not thousands, of years.

This also holds for Stephen Hawking. In the list of top 20 physicists in the 21st century shown in Table 4, Hawking ranks sixth and far ahead of some classic scientists including James Watt, Michael Faraday, and so on. In addition to Hawking's great scientific achievements, his passion in promoting science given his special physical condition has been seen as inspiring to the public and especially younger generations.

## 4. Conclusion and discussion

### 4.1. Summary

Big data is an important theme in natural and social sciences today. Adopting and expanding the approaches developed by Bohannon (2011a, 2011b), Michel et al. (2011) and Lin et al. (2012), this work exemplifies how to track the historical achievements and influence of great scientific minds. Utilizing the language option of Google Books Ngram Viewer, we find evidence in support of own-group preference, which explains the discrepancy of the UK survey and computational analysis

---

[7] The scientific fame of these 20 physicists persists after they pass away.

results based on the Google corpus. The co-occurrence analysis demonstrates the main contributions that scientists are known for by later generations

The wheel of history is always moving forward, with new developments emerging and old theories gradually fading away. This is an inexorable law of history, with each scientist's face and influence having a certain vitality cycle. Yet our study suggests the great minds live long intellectually. A scientist can only live several decades in physical form, but his contributions can live on in human history for hundreds or even thousands of years. Today when we talk about pioneers of science before the Christian era such as Aristotle, Euclid, and Archimedes, we find that these names have been written into history, respected and eulogized repeatedly by generations of newcomers. This may be the highest level of the meaning of life that scientists are pursuing.

### 4.2. Future research

It is an accepted practice to use citations to measure the scientific impact of researchers or their research output. Yet for an academic paper, the citation life cycle is usually about 3–5 years, and the count of citations often declines drastically afterwards (Glänzel & Schoepflin, 1995; Rogers, 2010; Tang, 2013). For a long time bibliometricians have used or abused this impact factor while also acknowledging the difficulties in quantifying long-term scientific impact (Waltman, 2016; Wang, Song, & Barabási, 2013). With the development of Google digitalized texts and computational analytical methods, we can track names being mentioned and assess the evolution of scientists' fame on a global scale over centuries. Given the variety of book types indexed in Google, this opens another possible research venue of altmetrics: measuring the influence of scientists beyond academia (Costas, Zahedi, & Wouters, 2015).

In his critical overview of altmetrics, Bornmann (2014) pointed out that no accepted framework existed to measure societal impact. Different from the existing altmetrics, which assess the impact of scientific work mostly on social media, we argue that when applied appropriately, Google Books can serve as a promising source of data for altmetrics, and the appearance frequency of scientists' names in books can serve as alternative metrics, which are still in great flux, to capture scientists' intellectual contributions to a broader public (Bornmann, 2014; Fenner, 2013; Piwowar, 2013). In the future it would be interesting to differentiate the fame of a scientist within academia and among the public with the categories provided in Google Books search (textbook, fiction, scholarly book, etc.). This paper focuses on physics; research looking into scientific fame of figures in other disciplines or even across disciplines is also worthy of further examination.

### 4.3. Limitations

In spite of their potential value for evaluation research, we would like to raise several notes of caution regarding the data and the method.

There are two sources of false positives. Searching individuals' names in books or scholarly publications can yield two types of results: works written by them and their achievements cited or mentioned by others. Given that Google Books and Ngram Viewer cannot differentiate the specific locations of search terms within books, false positives are inevitable. The good news is that even for an extremely prolific scientist, the actual number of articles one can write and publish is only in the hundreds. As an illustration of authorship versus citation, we searched Alexandre Dumas (1802–1870) who, though not a scientist, was a prolific French writer who dedicated his life to writing and published 357 books. We searched his full name in Google Books on December 20, 2017, and retrieved 23,400 hits. This suggests that the number of times mentioned can serve as a proxy indicator of fame of prominent individuals and is not a reflection of their authorship.

The other source of false positives is name disambiguation. For instance, searches for just "Newton" could result in mentions of other people or even the International System of Units (SI) unit of force. The common name problem, translation, and transliteration can also lead to an overestimation of scientists' fame. On the other hand, name changes over time associated with marriage and other reasons can produce false negatives, or an underestimation of reputation. In this sense the measurement of scientific fame works best for great scientists or scientists with unique names. For more details, please refer to Macroberts and Macroberts (1989) and Tang and Walsh (2010).

We also need to be aware of the issues of eponymy and obliteration, which have been discussed rather comprehensively by McCain (2011). When a scientist's achievements are so important as to become common knowledge, it is possible that only the achievements themselves rather than the scientist's full name are referenced. For instance, when people talk about the law of gravity, they may not mention Newton's name at all. In this case, using full name as a proxy causes an underestimation of fame. As McCain noted, "it is often observed that very well-established concepts, experimental methods, empirical findings, and models are mentioned in the literature without being linked to their originators" (p. 1413). This in our case will incur false negatives of recording scientists' fame. However, we argue that a scientist's full name never being mentioned once throughout a whole book is rare and thus has minimal impact on our findings.

It should be noted that for any computational analysis the amount of retrieved data should be sufficiently large. Scientific fame is no exception. Despite its potential, researchers should be aware of the caveats of the Google corpus, such as language bias against non-English languages, the quality heterogeneity of indexed books, and whether it is appropriate only for highly influential scientists rather than general scientists, when conducting fame analysis. It needs to be pointed out that though

Google Books and Ngram Viewer make it possible for us to analyze already famous scientists' fame, it would not be possible to identify those who are not yet famous scientists.[8]

Another caveat of this novel measurement is that it may not separate mentions of a famous figure for infamy rather than mentions of the figure for positive glory. Fortunately, in our case of two well-acknowledged great minds in science, the majority of the books on Newton and Einstein are based on a positive or at least a neutral evaluation. This saves us from the concern about evaluating controversial figures.

## Authors contribution

Guoyan Wang: Conceived and designed the analysis; Collected the data; Performed the analysis.
Guangyuan Hu: Wrote the paper.
Chuanfeng Li: Conceived and designed the analysis.
Li Tang: Conceived and designed the analysis; Contributed data or analysis tools; Wrote the paper.

## Acknowledgements

## References

Abdullah, A., & Thelwall, M. (2014). Can the impact of non-western academic books be measured? An investigation of google books and google scholar for Malaysia. *Journal of the Association for Information Science and Technology*, *65*(12), 2498–2508.
Acerbi, A., Lampos, V., Garnett, P., & Bentley, R. A. (2013). The expression of emotions in 20th century books. *PLoS One*, *8*(3), e59030.
Astin, A. V. (1957). Scientists and public responsibility. *Physics Today*, *10*(11), 23–27.
Baker, E. P. (1984). *Einstein versus Newton: The gravity of the situation*. Ohio: Online Computer Library Center.
Bohannon, J. (2011a). The science hall of fame. *Science*, *331*(6014), 143.
Bohannon, J. (2011b). Google books, Wikipedia, and the future of culturomics. *Science*, *331*(6014), 135.
Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, *8*(4), 895–903.
Bronk, D. W. (1976). Alfred Newton Richards (1876–1966). *Perspectives in Biology and Medicine*, *19*(3), 413.
Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019.
Egghe, L., & Rousseau, R. (2004). How to measure own-group preference? A novel approach to a sociometric problem. *Scientometrics*, *59*(2), 233–252.
Fagan, J. C. (2017). An evidence-based review of academic web search engines, 2014–2016: Implications for librarians' practice and research agenda. *Information Technology and Libraries*, *36*(2)
Feist, G. J. (2016). Intrinsic and extrinsic science: A dialectic of scientific fame. *Perspectives on Psychological Science*, *11*(6), 893–898.
Fenner, M. (2013). Letter from the guest content editor: Altmetrics have come of age. *Information Standards Quarterly*, *25*(2), 3.
Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on aging and reception processes of scientific literature. *Journal of Information Science*, *21*, 37–53.
Graneau, P., & Graneau, N. (1993). *Newton versus Einstein: How matter interacts with matter*. Sheffield: Carlton Press.
Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, *24*(9), 1722–1731.
Gribbin, J. (1987). Newton vs. Einstein. *The New Scientist*, *116*(1586), 45.
Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of google scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of Informetrics*, *11*(3), 823–834.
Hawking, S. W., & Israel, W. (Eds.). (1989). *Three hundred years of gravitation*. Cambridge University Press.
Hoover, E. R. (1977). *Cradle of greatness: National and world achievements of Ohio's Western Reserve*. Ohio: Cleveland.
Kelvin, L. (1901). Nineteenth century clouds over the dynamical theory of heat and light. *Philosophical Magazine*, *6*(2), 1–40.
Laubichler, M. D., Maienschein, J., & Renn, J. (2013). Computational perspectives in the history of science: To the memory of Peter Damerow. *ISIS*, *104*(1), 119–130.
Lin, Y., et al. (2012). Syntactic annotations for the Google Books Ngram Corpus. *ACL System Demonstrations*, *2*(1), 169–174.
Liu, W., et al. (2015). China's global growth in social science research. *Journal of Informetrics*, *9*(3), 555–569.
Liu, W., et al. (2018). Missing author address information in Web of Science-An explorative study. *Journal of Informetrics*, *12*(3), 985–997.
Luo, J. W. (2012). *The most influential 100 scientists in human civilization*. Xiamen: DAyi Publishing., in Chinese.
Macroberts, M. H., & Macroberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, *40*, 342–349.
McCain, K. W. (2011). Eponymy and obliteration by incorporation: The case of the "Nash equilibrium". *Journal of the American Society for Information Science and Technology*, *62*(7), 1412–1424.
Menard, H. W. (1971). *Science: Growth and change*. Cambridge, MA: Harvard University Press.
Merton, R. K. (1970). Behavior patterns of scientists. *American Scientist*, *3*(2), 213–220.
Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.
Newton, I. (1999). *The Principia: Mathematical principles of natural philosophy*. University of California Press.

---

Newton, I., Cohen, I. B., & Schofield, R. E. (1959). *Isaac Newton's papers and letters on natural philosophy*.

Pettit, M. (2016). Historical time in the age of big data: Cultural psychology, historical change, and the google books Ngram viewer. *History of Psychology*, *19*(2), 141–153.

Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, *493*(7431), 159.

Rogers, K. (2009). *The 100 Most Influential Scientists of All Time (The Britannica Guide to the World's Most Influential People)*. New York: Rosen Publishing.

Rogers, J. D. (2010). Citation analysis of nanotechnology at the field level: Implications of R&D evaluation. *Research Evaluation*, *19*(4), 281–290.

Roth, S. (2013). The fairly good economy: Testing the economization of society hypothesis against a Google Ngram View of trends in functional differentiation (1800–2000). *Journal of Applied Business Research*, *29*(5), 1495–1500.

Sanghera, P. (2011). *Quantum physics for scientists and technologists: Fundamental principles and applications for biologists, chemists, computer scientists, and nanotechnologists*. Wiley-Blackwell.

Simmons, J. G. (2009). *The 100 most influential scientists paperback*. New York: Fall River Press.

Sternfeld, J. (2011). Archival theory and digital historiography: Selection, search, and metadata as archival processes for assessing historical contextualization. *The American Archivist*, *74*(2), 544–575.

Stöver, D. J. H., & Trapp, J. (2007). *The life of Sir Charles Linnaeus*. Kessinger Publishing.

Tang, L. (2013). Does birds of a feather flock together̆matter-Evidence from a longitudinal study on US-China scientific collaboration. *Journal of Informetrics*, *7*(2), 330–344.

Tang, L., & Walsh, J. P. (2010). Bibliometric fingerprints: Name disambiguation based on approximate structure equivalence of cognitive maps. *Scientometrics*, *84*(3), 763–794.

Tang, L., Shapira, P., & Youtie, J. (2015). Is there a clubbing effect underlying Chinese research citation increases? *Journal of the Association for Information Science and Technology*, *66*(9), 1923–1932.

Tiner, J. H. (2002). *100 scientists who changed the world*. New York: Gareth Stevens Publishing.

Torre, D. L., Daleo, A., & García-Mata, &I. (2000). The Photon-Box Bohr-Einstein debate demythologized. *European Journal of Physics*, *21*(3), 253–260.

Twenge, J., Campbell, K. W., & Gentile, B. (2012). Increases in individualistic words and phrases in American books: I960–2008. *PLoS One*, *7*(7), e40181.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391.

Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127–132.

Whittaker, E. T. (1943). Aristotle, Newton, Einstein. *Science*, *98*(2542), 250–254.

Youtie, J., Rogers, J., Heinze, T., Shapira, P., & Tang, L. (2013). Career-based influences on scientific recognition in the United States and Europe: Longitudinal evidence from curriculum vitae data. *Research Policy*, *42*(8), 1341–1355.

Zeng, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to study implications of social and political change for cultural values. *International Journal of Psychology*, *50*(1), 47–55.