https://doi.org/10.51408/issi2025_092

Exploring Nobel Laureates' Question Selection Characteristics from a Topical Perspective*

Zou Xinran[1], Dong Yu[2], Wu Jiaxin[3]

[1]*zouxinran@mail.las.ac.cn*, [2]*dongy@mail.las.ac.cn*
National Science Library, Chinese Academy of Sciences, Beijing (China)
Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing (China)

[3]*shanganchong@163.com*
Department of Management Science and Engineering, School of Economics and Management, Wuhan University, Wuhan (China)

Abstract

Selecting a research question is the starting point of scientists' research activities, playing a crucial role not only in their career development, and further exploration is needed in this area. In this paper, we construct a characteristic index system for scientists' question selection based on the quality of attention theory in psychology, then conduct an empirical analysis using Nobel Laureates in natural sciences as examples, to reveal the characteristics of their question selection. Results show that Nobel laureates exhibit both commonalities and disciplinary differences in their question selection. Common characteristics include: a concentration of research topics in a limited number of directions, strong persistence in their research focus, and a balanced allocation of research effort between Broad and focus. Disciplinary differences are also evident. Physics laureates tend to engage in sustained and steady research across multiple interrelated fields. Chemistry laureates show a relatively higher degree of cross-disciplinary and cross-domain question selection; while they may moderately shift research directions over the course of their careers, these shifts typically revolve around one or a few core themes. In contrast, laureates in Physiology or Medicine display more exploratory question selection behaviors, frequently switching among one or several related core areas, with comparatively lower research persistence. Across different stages of their careers, the three groups of laureates demonstrate distinct question selection patterns. Physics laureates tend to broaden the scope of their research while simultaneously deepening their focus. Chemistry and Physiology or Medicine laureates follow a similar trajectory characterized by early-stage broad exploration, mid-career flexibility, and late-career deep focus. These findings highlight the varying research patterns across disciplines and offer valuable insights into how Nobel laureates select and shift their research questions.

Introduction

Selecting a research question is the starting point of scientists' research activities, playing a crucial role not only in their career development but also in shaping the progress of their discipline (Ding et al., 2023; Yu et al., 2021; Foster et al., 2015).

* This work is supported by the Major Project of the National Social Science Foundation of China, “Research on Talent Training Models and Strategic Deployment in Key and Core Technology Fields in China” (22&ZD127).

Exploring the characteristics of research question selection provides valuable insights into scientists' career development process and contributes to a deeper understanding of their research patterns. Moreover, it holds important practical implications for science and technology management, particularly in areas such as talent cultivation, research funding allocation, and disciplinary development.

The selection of research questions by scientists has long been a central topic of interest in the fields of sociology and scientometrics (Ding et al., 2023; Van Houten et al., 1983). As early as the 1980s, scholars began to investigate field mobility among physicists. Early studies were primarily based on qualitative research and simple questionnaire surveys. In recent years, the development of large-scale bibliometric datasets—such as Scopus, Microsoft Academic Graph, and the American Physical Society (APS) dataset—has provided rich data sources for quantitatively analyzing scientists' question selection. Moreover, the emergence of advanced techniques such as natural language processing and complex network analysis has offered new perspectives for exploring scientists' research behavior in greater depth (Liu et al., 2023; Taylor et al., 2022).

Current research on scientists' question selection has produced rich findings, with a primary focus on their question selection performance. However, the construction of measurement indicators related to the question selection performance often lacks theoretical grounding, resulting in a degree of subjectivity and the absence of a comprehensive and objective evaluation framework. Based on this, we introduce the attention theory in psychology and construct a characteristic index system of scientists' question selection. Then we conduct an empirical analysis using Nobel Laureates in natural sciences as examples. This study aims to enhance existing research and provide a more scientific basis and reference for the career development, talent training, and research funding policies of Chinese scientists.

Literature review

Analysis process and method of scientists' question selection characteristics

Existing studies primarily reveal the question selection characteristics of scientists by measuring the topic transition in their careers. The measurement process follows three main steps: (1) identifying research topics, (2) dividing time periods, and (3) measuring topic shifts.

There are three ways to identify the research topics of scientists. Many studies obtain the topic vector of a paper by vectorizing the topic code or research field assigned to the paper by the database. For example, the PACS code in the American Physical Society (APS) dataset consists of six letters and numbers, in which the first two numbers define 67 major topics in the field, and the topic vectors of multiple papers can be obtained by statistically normalizing the frequency of the first two numbers (Jia et al., 2017). The second is to construct citation or co-occurrence networks of papers and perform community clustering so as to classify each paper under different subject categories. One study uses papers as nodes to build a paper co-citation network for each scientist. That is, if two papers cite the same references, they are linked together, and then use the Fast Unfolding algorithm to identify each scientist's co-citation network. The primary communities identified represent the scientist's main research topics (Zeng et al., 2019). Thirdly, the research topics can be identified from the title, keywords, abstract and other text information of the paper. For example, Bert (Ding & Chen, 2023), Top2Vec (Chen et al., 2019) are used to vectorize the text information such as the title of the paper, obtain the topic vector of each paper, and further determine the topic category of the paper through clustering. The evolution of a scientist's research topics is a dynamic process, and analyzing their entire career as a whole may overlook important temporal variations. To address this, existing studies segment a scientist's career into different time stages based on three main approaches: The fixed time interval approach groups papers published within a set number of consecutive years into the same stage. For instance, papers published in 2000 and 2001 are classified as one stage, while those from 2001 and 2002 form the next, and so on (Huang et al., 2023). The publication count approach segments a scientist's career by grouping a fixed number of consecutive papers into the same stage, regardless of the publication year. For example, some studies define each stage as a block of m consecutive papers (Huang et al., 2023). The key career milestone approach divides a scientist’s career based on significant events such as the publication of their most cited paper or receiving an award. For instance, some researchers classify Nobel laureates' careers into three stages: before publishing their prize-winning paper, after publishing it, and post-award (Ding & Chen, 2023).

Existing research measures the characteristics of scientists' question selection across several dimensions: topic transition speed, topic transition span, topic focus intensity, and topic coverage. Topic transition speed measures how quickly scientists shift between different topics, reflecting their level of concentration on specific research problems. It can be quantified by calculating the amplitude of topic change over time or by the ratio of topics to papers within a given period (Chen et al., 2019). Topic transition span assesses the degree of content difference before and after a topic shift, indicating whether scientists tend to explore significantly new research directions. This is typically measured using cosine similarity (Liu et al., 2024) or Euclidean distance (Liu & Xia, 2017). Topic focus intensity captures the level of attention scientists devote to a particular topic within a specific timeframe, revealing whether they prefer deep exploration of a single issue. It is commonly measured by the number of papers published on a given topic—where a higher count indicates greater investment (Ding & Chen, 2023), or by the proportion of papers on a topic relative to the scientist's total output (Chen et al. 2023; Chen et al., 2019). Topic coverage reflects the breadth and diversity of a scientist’s research within a given period. It can be measured by the number of topics studied (Ding & Chen, 2023), or spatially by calculating the coverage area of topic vectors. For example, the volume of the smallest ellipsoid encompassing all topic vectors can serve as a proxy (Bu et al., 2022).

Analysis subjects and conclusions of scientists' question selection characteristics

Most researches primarily focuses on scientists in physics and computer science. Studies on physicists often use the APS dataset, which assigns PACS codes to papers for topic identification. Research on computer scientists relies on multiple databases, including Microsoft Academic Graph (MAG), Microsoft Academic Search, and DBLP, all of which classify papers by research field. From a group perspective, some studies analyze a single cohort, such as all physicists or Nobel laureates in physics, while others compare multiple groups to examine differences in question selection。 Findings reveal some differences between elite scientists and the broader scientific. community. In physics, general scientists tend to expand their research over time: increasing the speed (Zeng et al., 2019), span (Aleta et al., 2019) and topic coverage (Zeng et al., 2019). In contrast, Nobel laureates remain more focused, dedicating long-term attention to their prize-winning topics and later expanding on related topics (Ding & Chen, 2023). A similar pattern is observed in computer science, where high-impact, high-productivity scientists exhibit greater research focus compared to others (Liu et al., 2024). Additionally, question selection characteristics vary across disciplines. For instance, scientists in physics are more likely to work on multiple topics simultaneously in the middle of their careers (Zeng et al., 2019), while scientists in computing are more likely to work early and late in their careers (Chakraborty et al., 2015). Although differences in research methodology and indicator selection may limit cross-disciplinary comparisons, the findings still highlight that scientists' question selection characteristics vary across disciplines.

Research Design

Topic identification

To measure the characteristic of scientists' question selection, it is essential to first identify the research topics of each scientist. We extract the research topics of each scientist from the collected paper datasets. Through text embedding, dimensionality reduction and topic clustering, the number of research topics for each scientist, the topic vectors of the research topics, and the topic classification of each paper are obtained.

Text Embedding

The title, abstract, and keywords of a paper condense the main research content of the paper and are a refined representation of the paper's topics. Text embedding can map the title, abstract, keywords and other information of the paper into vectors in the space, and extract the topic vector of the paper as the basis for subsequent topic clustering. SciBERT is a pre-trained language model based on the BERT architecture. Trained on a large-scale corpus of scientific publications, SciBERT offers stronger language understanding and semantic representation capabilities for scientific texts compared to other pre-trained models. Therefore, we use SciBERT to extract the topic vector of each paper.

Dimensionality reduction

In high-dimensional space, the distances among samples can become strikingly similar, making it difficult for clustering algorithms to distinguish distinct data characteristics. To alleviate this problem, we use the UMAP to reduce the dimensions of the vectors output by SciBERT.

Topic Clustering

AP (Affinity Propagation) clustering is suitable for small to medium-sized datasets and does not require a predefined number of clusters. Since most authors have fewer than 800 publications, this study applies the AP clustering algorithm to cluster the topic vectors of each scientist's papers. The clustering process yields the number of research topics for each scientist, assigns a topic label to each paper, and computes the average vector of papers within each cluster as the representation of that topic. The clustering results are evaluated using the Silhouette Coefficient (SC), which ranges from -1 to 1, with higher values indicating better clustering quality. Based on the SC, the number of research topics for each scientist is determined.

Research theory and Index construction

Psychological research widely recognizes that attention consists of four fundamental dimensions (Meng, 1994): attentional span, referring to the number of objects one can focus on simultaneously; attentional stability, denoting the ability to sustain attention on a specific perception or activity over time; attentional allocation, indicating the capacity to distribute attention across multiple objects or activities at the same time; and attentional shifting, describing the active, purposeful, and timely transition of attention from one object or activity to another.

Scientist's research question selection can be regarded as the allocation of research attention across different topics. Since a scientist's research attention is limited, they may adopt different allocation strategies, resulting in diverse patterns of question selection. Based on this, we refer to the attention quality theory described above to constructs an index system from four dimensions: span, stability, distribution and transfer.

Span dimension

Scientists may engage in multiple research topics within a given period. The degree of content variation among these topics reflects the breadth and diversity of their question selection. Accordingly, **Topic Coverage Index (*C*)** is introduced to measure the extent of content differentiation across a scientist's research topics during a specific time period. The formula is as shown in Eq. (1):

$$C = 1 - \min_{1 \le i < j \le N} s_{i,j} \tag{1}$$

$s_{i,j}$ represents the cosine similarity between v_i and v_j, calculated as shown in Eq. (2):

$$s_{i,j} = \frac{v_i \cdot v_j}{\|v_i\| \cdot \|v_j\|} \tag{2}$$

Where, v represents the topic vector of a scientists, and N represents the total number of research topics of the scientist.

Allocation dimension

When scientists engage in multiple research topics simultaneously, the amount of research effort devoted to each topic may vary. This can be measured by the number of publications under each topic—more publications in a given topic indicate a greater allocation of research effort to that area. An uneven distribution of publications across topics suggests the presence of core research areas, while a more balanced distribution indicates that the scientist tends to allocate research efforts more evenly across topics. Based on this, the **Topic Focus Index (F)** is introduced to reflect the degree of evenness in a scientist's allocation of research effort across different topics. The calculation is based on Pielou's Evenness Index, as shown in Eq. (3):

$$F = \frac{H'}{\log(N)} \quad (3)$$

H' is Shannon entropy, which can be calculated as shown in Eq. (4):

$$H' = -\sum_{i=1}^{N} p_i \log(p_i) \quad (4)$$

Where, p_i is the proportion of the number of papers under the i topic to the total number of papers, and the calculation formula is in Eq. (5)

$$p_i = \frac{n_i}{n} \quad (5)$$

The value of F ranges from 0 to 1, where a value closer to 1 indicates a more even distribution of a scientist's papers across topics. Here, n_i represents the number of papers published under topic i, n denotes the total number of papers published by a scientist, and N represents the total number of research topics explored by the scientist.

Stability dimension

Scientists may continue working on the same research topic over an extended period. The duration of sustained engagement with a topic reflects the persistence and stability of their research. The **Topic Duration Index (G)** is introduced to measure the proportion of a scientist's career spent on each topic, averaged across all topics. The formula is in Eq. (6):

$$G = \frac{1}{T \cdot N} \sum_{i=1}^{N} (t_i + 1) \quad (6)$$

Where t_i represents the difference between the publication year of the last and first paper under the topic i, N represents the total number of topics studied by the scientist, and T represents the span of a scientist's research career.

Shifting dimension

Scientists may shift research topics throughout their careers. The **Topic Shifting Speed Index (S)** is introduced to measure the frequency of transitions between different research topics over time. The formula is in Eq. (7):

$$S = \frac{1}{T-1} \sum_{i=2}^{T} N_i \quad (7)$$

Where T represents the number of years in which a scientist has published papers, i denotes a specific publication year, and N_i represents the number of new topics introduced in year i compared to year i-1. For example, if a scientist studied topic1, topic2, topic3 in year i-1, and in year i studied topic 1, topic2, topic4 and topic5, then N_i would be 2.

Empirical Study

Figure 1 shows the technology roadmap.

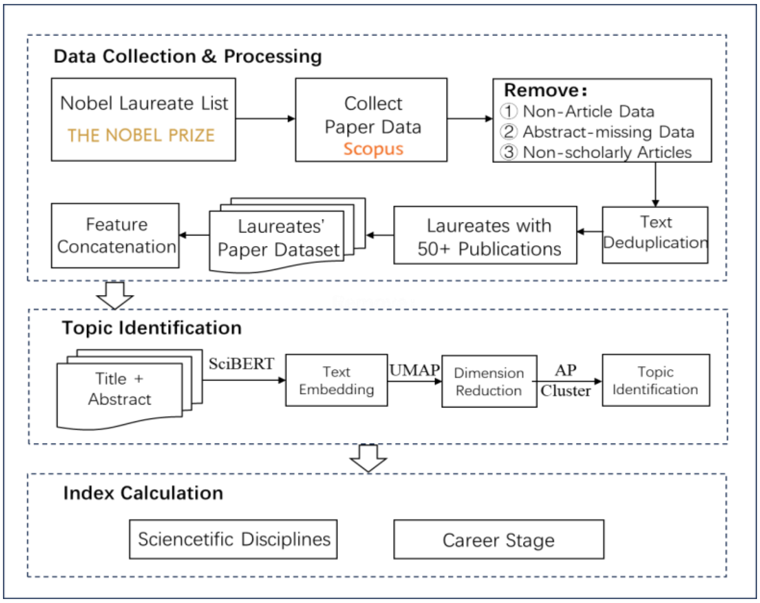

Figure 1. Technology Roadmap.

Data Collection and Cleaning

The Nobel Prize is one of the highest honors in science, awarded to individuals who have made "the greatest benefit to humankind." As elite scientists in different fields, analyzing the characteristics of Nobel laureates can provide valuable insights for the career development of young researchers. Therefore, we select Nobel Laureates from 1901 to 2023 as the research subjects.

Data Collection

First, collect the basic information of Nobel laureates from the Nobel Prize website, including name, award year, etc. Then collect and clean the publication of Nobel laureates. According to the evaluation of author identification effect in WOS, Scopus, AMiner, OpenAlex, and ORCID by Shi Dongbo et al. (2024), Scopus outperforms others in coverage, accuracy, and robustness. Therefore, we use Scopus as the primary data source. Each laureate's personal page is accessed using their Scopus ID, and a dataset of their published papers (retrieved in June 2024) is downloaded, including the title, abstract, keywords, publication date and so on.

Data Cleaning

The collected data is further cleaned as follows: (1) only records classified as “Article” are retained; (2) duplicate entries are removed; (3) records lacking abstract information are excluded; and (4) non-research articles such as Nobel Lectures are removed. Considering that the publication records of some early laureates—especially those awarded before the mid-20th century—may be incomplete, which could affect analysis accuracy, we include only laureates with at least 50 publications. The final dataset comprises 366 Nobel laureates and a total of 82,879 papers, including 123 laureates in Physics (24,116 papers), 123 laureates in Chemistry (32,586 papers), and 120 laureates in Physiology or Medicine (26,177 papers).

Results

Topic Identification Analysis

The distribution of the number of research topics is shown in Figure 2. It can be seen that the number of laureates with 2 research topics in their entire career is the largest, and more than 85% of the laureates have 2 to 5 research topics.

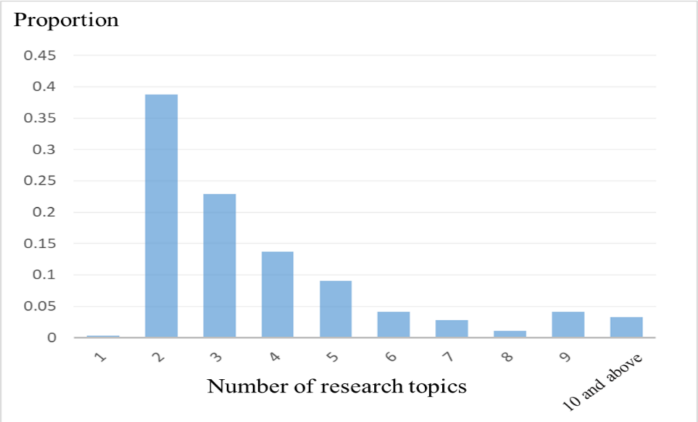

Figure 2. Distribution of the Number of Research Topics of All Laureates.

The average number of research topics of Nobel laureates in different disciplines is shown in Table 1. It can be seen that the average number of research topics of Nobel laureates in Physiology or Medicine in their entire career is the highest, which is 4.2 topics, higher than the average number in chemistry (3.7 topics) and physics (3.4 topics).

Table 1. Average Number of Research Topics of Scientists.

Award Category	*Mean*
Physics	3.4
Chemistry	3.7
Physiology or Medicine	4.2

Index Characteristic Analysis

Each Nobel laureate is calculated separately according to the index calculation method. Then calculate the average to represent the average level of each disciplinary field.

Topic Coverage Index (C)

Figure 3 and Table 2 respectively illustrate the distribution of topic coverage and related statistical indicators for laureates in Physics, Chemistry, and Physiology or Medicine. In terms of average values, the topic coverage among laureates across these disciplines is relatively close. Chemistry laureates show a slightly higher average topic coverage (0.014) compared to laureates in Physics (0.013) and Physiology or Medicine (0.013), suggesting that, overall, Chemistry laureates tend to work on topics with greater differences in content, indicating more diversity and interdisciplinarity in their research choices. In terms of standard deviation, Physics laureates exhibit the highest variability (0.023), significantly higher than that of laureates in Physiology or Medicine (0.085) and Chemistry (0.047), indicating that Physics laureates show the greatest internal differences in topic coverage within their group.

A closer look at the distribution of topic coverage reveals a right-skewed pattern in all three disciplines, with most values concentrated in the 0–0.02 range. This suggests that most laureates tend to focus on topics with relatively small differences throughout their research careers, concentrating on a limited number of directions. From the perspective of disciplinary differences, Physics laureates exhibit a longer distribution tail, with a maximum value reaching 0.148—substantially higher than the maximums for Chemistry (0.047) and Physiology or Medicine (0.085). This indicates the presence of a small number of Physics laureates whose research spans exceptionally broad areas. One such example is Rainer Weiss, who ranked second in topic coverage and was awarded the 2017 Nobel Prize in Physics for the development of the LIGO detector and the observation of gravitational waves. Over the course of his career, he worked on nine different topics, ranging from particle physics and astrophysics to gravitational wave astronomy, clearly demonstrating strong interdisciplinarity. In Chemistry and Physiology or Medicine, most laureates exhibit lower topic coverage values. However, a small "secondary peak" appears around 0.04, indicating that in these two fields, there is a subset of laureates whose research topics are relatively diverse—though not to the same extent as those in Physics.

Table 2. Descriptive Statistics of Topic Coverage Index (C).

Statistical Indicator	*Physics Laureates*	*Chemistry Laureates*	*Physiology or Medicine Laureates*
Mean	0.013	0.014	0.013
Minimum	0.000	0.000	0.000
Upper Quartile	0.003	0.004	0.004
Median	0.006	0.008	0.007
Lower Quartile	0.014	0.017	0.013
Maximum	0.148	0.047	0.085
Standard Deviation	0.023	0.015	0.016

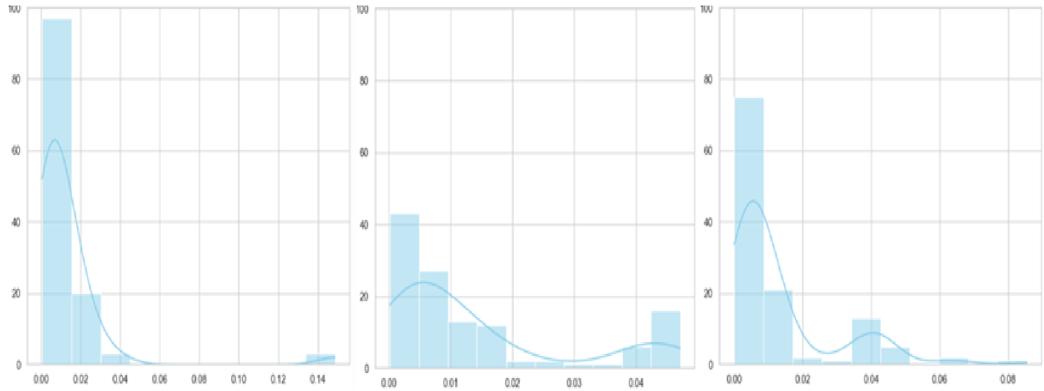

Figure 3. Topic Coverage Index (C) distribution (physics vs. chemistry vs. Physiology or medicine).

Topic Focus Index (F)

As shown in Figure 4 and Table 3, the three disciplinary groups of Nobel laureates exhibit relatively similar patterns in topic concentration. Comparatively, Physics laureates have a slightly higher average topic concentration (0.63) than Chemistry (0.59) and Physiology or Medicine laureates (0.59), indicating that Physics laureates tend to distribute their research efforts more evenly across topics throughout their careers, while Chemistry and Physiology or Medicine laureates tend to focus their research efforts more narrowly. In terms of standard deviation, the topic concentration of Physiology or Medicine laureates shows the highest variation (0.25), slightly higher than that of Physics (0.24) and Chemistry laureates (0.23), suggesting greater within-group differences in how these laureates allocate their research efforts across topics. However, this difference is not particularly significant compared with the other two disciplines.

A closer look at the distribution of topic concentration reveals a general left-skewed trend across all three disciplines, with peaks concentrated in the 0.6–0.8 range. The maximum values of topic concentration in all three disciplines are close to 1 (rounded), indicating that in each field, there are laureates who distribute their research efforts almost equally among multiple topics. For example, Koichi Tanaka, who won the Nobel Prize in Chemistry in 2002, distributed his research efforts nearly evenly across two topics over his career. Topic 1 involved the synthesis and reaction mechanisms of small organic molecules, including cycloaddition and aromatic substitution reactions, with 97 papers. Topic 2 focused on the structure and electronic properties of large conjugated systems, under which he published 95 papers.

Table 3. Descriptive Statistics of Topic Focus Index (F).

Statistical Indicator	*Physics Laureates*	*Chemistry Laureates*	*Physiology or Medicine Laureates*
Mean	0.63	0.59	0.59
Minimum	0.05	0.07	0.00
Upper Quartile	0.51	0.45	0.40
Median	0.66	0.63	0.64
Lower Quartile	0.80	0.76	0.79
Maximum	1.00	1.00	1.00
Standard Deviation	0.24	0.23	0.25

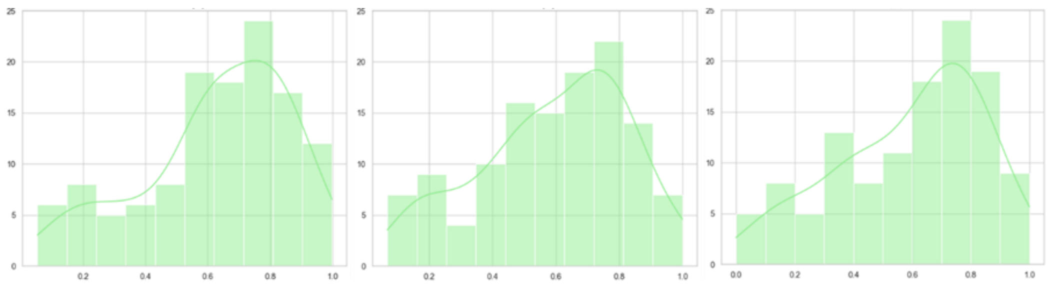

Figure 4. Topic Focus Index (F) distribution (physics vs. chemistry vs. Physiology or medicine).

Topic Duration Index (G)

Figure 5 illustrates the distribution of topic persistence among Nobel laureates in the three disciplines, while Table 4 presents the corresponding descriptive statistics. In terms of average values, Physics laureates have a topic persistence of 0.72, Chemistry laureates 0.71, both higher than that of Physiology or Medicine laureates, which stands at 0.66. This indicates that Nobel laureates generally demonstrate strong research persistence, with Physics and Chemistry laureates showing a particularly prominent tendency to pursue long-term research on a single topic. Regarding standard deviation, Physiology or Medicine and Physics laureates both have a topic persistence standard deviation of 0.17, slightly higher than that of Chemistry laureates (0.14), suggesting that the former two groups exhibit greater internal variability in their research persistence. Chemistry laureates, in contrast, display more consistent persistence overall.

Further analysis of the distribution reveals that the topic persistence of Physics laureates is mostly concentrated above 0.5, with a notable peak around 0.85 and a considerable number of laureates approaching a persistence score of 1. Overall, Physics laureates exhibit high topic persistence, with over 75% of them dedicating more than half of their career to a single topic—demonstrating a strong commitment to long-term research. For instance, Horst L. Stormer, who won the Nobel Prize in Physics in 1998, had a topic persistence as high as 0.99. He published his first paper in 1976 and had a research career spanning 35 years. During this period, he focused on two main topics: (1) semiconductor materials and electronic transport, including modulation doping and two-dimensional electron gases, and (2) his well-known research on the quantum Hall effect and its physical mechanisms. His research spanned 34 and 35 years on these two topics respectively, exemplifying deep and continuous engagement in specific research areas. Chemistry laureates show a more symmetrical distribution of topic persistence, primarily ranging from 0.5 to 0.9, with a relatively flat peak, suggesting a balanced overall pattern. There are fewer laureates at the extreme low or high ends of the scale. In contrast, the topic persistence of Physiology or Medicine laureates displays a bimodal distribution, with two main peaks: one between 0.5 and 0.6, and another between 0.75 and 0.85. The peak around 0.5 is the highest, indicating that in this field, the largest group of laureates falls into the moderate range of topic persistence—on average, spending about half of their research careers focused on a single topic.

Table 4. Descriptive Statistics of Topic Duration Index (G).

Statistical Indicator	*Physics Laureates*	*Chemistry Laureates*	*Physiology or Medicine Laureates*
Mean	0.72	0.71	0.66
Minimum	0.28	0.37	0.28
Upper Quartile	0.60	0.59	0.52
Median	0.73	0.72	0.65
Lower Quartile	0.86	0.83	0.80
Maximum	0.99	0.99	1.00
Standard Deviation	0.17	0.14	0.17

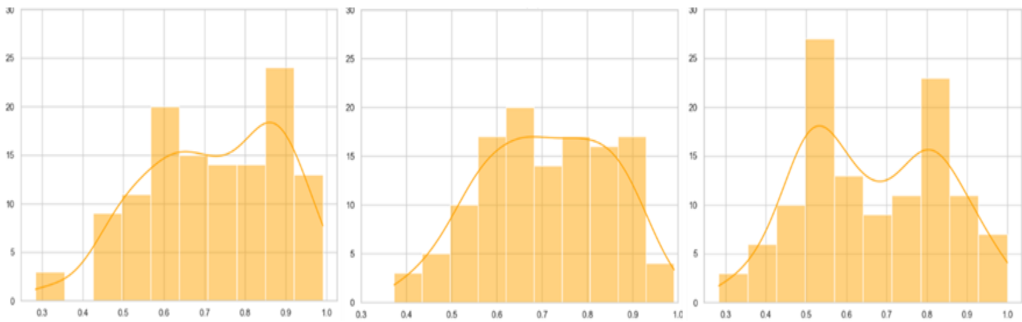

Figure 5. Topic Duration Index (G) distribution (physics vs. chemistry vs. Physiology or medicine).

Topic Shifting Speed

Figure 6 and Table 5 respectively present the distribution and descriptive statistics of topic switching speed among Nobel laureates in Physics, Chemistry, and Physiology or Medicine. In terms of average values, laureates in Physiology or Medicine have the highest topic switching speed at 0.40, followed by Chemistry laureates at 0.37 and Physics laureates at 0.35. This indicates that Physiology or Medicine laureates tend to switch between research topics more frequently throughout their careers, continuously exploring directions different from their current research. In terms of standard deviation, Chemistry laureates exhibit the highest variation in topic switching speed (0.32), suggesting substantial internal differences in how frequently they change research topics. In contrast, Physics laureates show the lowest standard deviation (0.25), indicating a more consistent pattern across the group, with generally lower switching speeds.

Further analysis of the distribution shows that topic switching speed in all three disciplines is significantly right-skewed, with most laureates' switching speeds concentrated between 0.2 and 0.4—particularly in Chemistry. This suggests that most laureates have relatively low switching speeds over the course of their careers, tending to stay focused on their current lines of research. However, the disciplines differ more notably at the extremes. The maximum switching speed among Chemistry laureates reaches as high as 1.62, which is higher than that of Physiology or Medicine laureates (1.33) and Physics laureates (1.26). This implies that a small number of Chemistry laureates change research topics extremely frequently. One example is Roald Hoffmann, a Chemistry Nobel laureate with a switching speed of 1.62. Over the course of his career, he explored 12 different research topics, showing a clear pattern of shifting directions. He began with the development of quantum chemistry and molecular orbital theory, then applied these theoretical methods to the analysis of organic reaction mechanisms. His research later expanded into the structural and reactive properties of inorganic and organometallic compounds, and ultimately extended to the theoretical design of electronic structures in solid-state materials and novel conductive systems.

Table 5. Descriptive Statistics of Topic Shifting Speed (S).

Statistical Indicator	*Physics Laureates*	*Chemistry Laureates*	*Physiology or Medicine Laureates*
Mean	0.35	0.37	0.40
Minimum	0.03	0.04	0.00
Upper Quartile	0.18	0.16	0.18
Median	0.30	0.27	0.33
Lower Quartile	0.44	0.49	0.56
Maximum	1.26	1.62	1.33
Standard Deviation	0.25	0.32	0.29

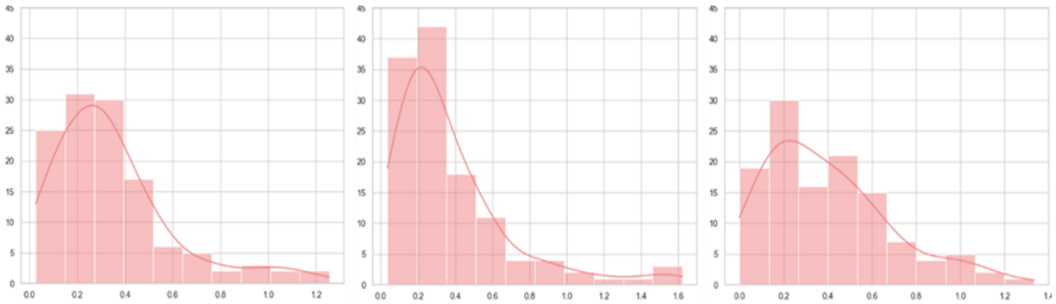

Figure 6. Topic Shifting Speed Index (S) distribution (physics vs. chemistry vs. Physiology or medicine).

Trends in characteristic indicators across different career stages

Figure 7 illustrates the changing trends of characteristic indicators across different career stages of Nobel laureates in Physics, Chemistry, and Physiology or Medicine. It can be seen that, except for the Topic Coverage Index (C), the other three indicators show similar patterns of change.

Figure 7. Trends in characteristic indicators across different career stages.

From the trend in topic coverage, Physics laureates exhibit a distinctive upward trajectory as their careers progress, indicating that the diversity of their research content increases over time. This suggests enhanced interdisciplinarity and greater variation in their question selection. In contrast, Chemistry and Physiology or Medicine laureates show a declining trend in topic coverage, reflecting a reduction in content diversity and a growing tendency toward thematic convergence.

With respect to topic concentration, all three disciplines demonstrate a gradual downward trend across the career span, with a particularly pronounced decline during the late career stage. This pattern reflects a transition from a relatively dispersed allocation of research efforts in the early career stage to a more focused investment in core topics over time, highlighting an increasing degree of specialization. This also implies that Nobel laureates tend to explore multiple fields in the early stages of their careers, but as their research interests become more defined, they increasingly build upon and deepen their existing research foundations. Notably, Physics laureates consistently maintain a more evenly distributed research effort—not only across their entire career trajectories but also within the early, mid, and late career stages.

In terms of topic persistence, an overall increasing trend is observed across career stages. Physics laureates demonstrate relatively high persistence from the early career stage, which continues to rise steadily as their careers advance. Chemistry and Physiology or Medicine laureates show similar trajectories, with a substantial increase in persistence during the mid-career stage compared to the early stage, followed by a relatively stable level thereafter. This suggests that long-term engagement with specific research topics becomes increasingly prominent as laureates progress through their careers.

Regarding topic switching speed, laureates across all three disciplines tend to exhibit the highest switching rates during the mid-career stage, indicating a greater inclination to explore new directions distinct from their current research. Notably, Physiology or Medicine laureates demonstrate a consistently high switching tendency not only throughout the entire career span but also within the early, mid, and late stages, suggesting a more dynamic and exploratory research pattern within this field.

Conclusion and Discussion, Future work

In this paper, we construct an index system for scientists' question selection and conduct an empirical analysis using Nobel Laureates in natural sciences as examples. This study unveils the multi-dimensional characteristics of scientists' research question selection, offering insights into the research patterns of scientists' question selection.

In conclusion, Nobel laureates share several common characteristics in their research question selection. First, their research topics are typically concentrated in a limited number of directions, with a certain degree of thematic relatedness. Second, they exhibit strong research persistence, often engaging in long-term, in-depth exploration of a single topic. Third, they tend to strike a balance between focused and dispersed allocation of research efforts. At the same time, there are notable disciplinary differences in question selection characteristics. Compared to other fields, Physics laureates show stronger thematic coherence, more evenly distributed research efforts, longer durations of topic engagement, and lower switching frequency. Their question selection is characterized by sustained and stable advancement across multiple interrelated research areas. Chemistry laureates, by contrast, display greater interdisciplinarity and cross-domain exploration. Although they may switch research directions during their careers, such changes typically revolve around one or a few core themes. Physiology or Medicine laureates exhibit more exploratory question selection patterns, frequently shifting between one or several related core topics, with relatively lower research persistence.

As their careers progress, Physics laureates demonstrate a tendency to both broaden the scope and deepen the focus of their research. This is manifested in increasing interdisciplinarity and diversity in question selection, a gradual concentration of research efforts, and steadily strengthening research persistence. Chemistry and Physiology or Medicine laureates, on the other hand, generally follow a similar trajectory characterized by broad exploration in the early career stage, flexible adjustment in the mid-career stage, and focused deepening in the late career stage. This pattern is reflected in a gradual reduction in the diversity of research content, the emergence of core research themes, and a progressive increase in research persistence, which remains relatively stable in the later stages of their careers. These findings not only shed light on the nuanced evolution of question selection among Nobel laureates across disciplines but also provide valuable insights into the dynamic interplay between research breadth and depth throughout a scientific career.

However, there are still limitations. First, the analysis requires further interpretation and robustness verification. Future work will incorporate qualitative validation through interviews, biographies, and other textual data from scientists at different career stages to enhance the credibility of the results. Second, the focus on Nobel laureates limits the analysis, excluding comparisons with scientists at other levels. Future research will include data from scientists at various levels, such as members of the American Academy of Sciences and other researchers, for a comparative analysis.

References

Aleta, A., Meloni, S., Perra, N., & Moreno, Y. (2019). Explore with caution: mapping the evolution of scientific interest in physics. EPJ Data Science, 8(1). doi:10.1140/epjds/s13688-019-0205-9

Bu, Y., Huang, S., Huang, Y., & Lu, W. (2022). Temporal-Spatial Measurements for Research Topic Evolution of Researchers: Speed, Volume, and Circuitousness. Library and information service, 66(24), 84-91.

Chakraborty, T., Tammana, V., Ganguly, N., & Mukherjee, A. (2015). Understanding and modeling diverse scientific careers of researchers. Journal of Informetrics, 9(1), 69-78. doi:10.1016/j.joi.2014.11.008

Chen, L., Guo, S., Teng, G., & Tuo, R. (2019). Research Topic Focus and Transfer of Scientific Personnel. Digital Library Forum(12), 9-17.

Chen, X., Pan, Y., Ma, Z., Zhang, G., Zhang, B., & Ren, Q. (2023). The Influence of Fund Support on the Research Direction of Outstanding Young Scholars. Journal of the China Society for Scientific and Technical Information, 42(12), 1438-1447.

Ding, J., & Chen, Y. (2023). Concentration, continuation and Extension: Research Pattern of Nobel Laureates Based on the Topic: An Empirical Analysis of Physics. Library Journal, 42(08), 100-109.

Ding, J., Chen, Y., & Liu, C. (2023). Exploring the research features of Nobel laureates in Physics based on the semantic similarity measurement. Scientometrics, 128(9), 5247-5275. doi:10.1007/s11192-023-04786-3

Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and Innovation in Scientists' Research Strategies. American Sociological Review, 80(5), 875-908. doi:10.1177/0003122415601618

Huang, S., Huang, Y., Bu, Y., Luo, Z., & Lu, W. (2023). Disclosing the interactive mechanism behind scientists' topic selection behavior from the perspective of the productivity and the impact. Journal of Informetrics, 17(2). doi:10.1016/j.joi.2023.101409

Jia, T., Wang, D., & Szymanski, B. K. (2017). Quantifying patterns of research-interest

evolution. Nature Human Behaviour, 1(4). doi:10.1038/s41562-017-0078
Liu, F., & Xia, H. (2017). Discovering Interest Transition Patterns of Technological Inventors. Technology Intelligence Engineering, 3(02), 33-40.
Liu, L., Jones, B. F., Uzzi, B., & Wang, D. S. (2023). Data, measurement and empirical methods in the science of science. Nature Human Behaviour, 7(7), 1046-1058. doi:10.1038/s41562-023-
Liu, M., Shi, J., Yang, S., & Bu, Y. (2024). Speed of Research Topic Evolution and Scientific Performance: Evidence from Computer Science. Library and information service, 68(06), 72-82.
Meng, Z. (1994). General psychology. Beijing: Peking University Press.
Shi, D., Deng, H., Yang, Z., Liu, N., Liu, Y., & Mao, Y. (2024). A Study on the Applicability of Author Identification Numbers in Scientific and Technical Paper Databases. ChinaXiv. doi:doi:10.12074/202406.00022V2
Taylor, J. A., Larraondo, P., & de Supinski, B. R. (2022). Data-driven global weather predictions at high resolutions. INTERNATIONAL JOURNAL OF HIGH PERFORMANCE COMPUTING APPLICATIONS, 36(2), 130-140. doi:10.1177/10943420211039818
Van Houten, J., Van Vuren, H., Le Pairs, C., & Dijkhuis, G. (1983). Migration of physicists to other academic disciplines: situation in the Netherlands. Scientometrics, 5(4), 257-267.
Yu, X., Szymanski, B. K., & Jia, T. (2021). Become a better you: Correlation between the change of research direction and the change of scientific performance. Journal of Informetrics, 15(3). doi:10.1016/j.joi.2021.101193
Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., . . . Havlin, S. (2019). Increasing trend of scientists to switch between topics. Nat Commun, 10(1), 3439. doi:10.1038/s41467-019-11401-8