Information Processing and Management 62 (2025) 104171



Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman



Unraveling topic switching and innovation in science

Alex J. Yang *

School of Information Management, Nanjing University, Nanjing 210023, PR China

ARTICLE INFO

Keywords:
Science of science
Scientific career
Topic switch behavior
Disruption
Novelty

ABSTRACT

The selection of research topics shapes both individual scientific trajectories and the broader evolution of knowledge. Despite its critical role, a systematic investigation into the dynamics of topic switching among scientists and its relationship with scientific innovation remains limited. Drawing on a comprehensive dataset encompassing the career trajectories of 1.4 million scientists and 27.6 million publications from 1950 to 2020, I use a field-free and finely-grained framework to quantify shifts in research direction by measuring the knowledge distance between a paper's references and those of prior works. To account for systemic biases, I construct a null model that captures expected patterns of topic selection. My analysis reveals three key findings: (1) Scientists exhibit lower-than-expected levels of topic switching, with a decline before 2000 followed by a rising trend thereafter; (2) Early-career researchers, female scientists, and non-elite scientists demonstrate higher levels of topic switching compared to their counterparts; and (3) Increased topic switching correlates with greater research novelty, interdisciplinarity, and disruptive potential. These findings provide valuable insights into the mechanisms underlying scientific exploration and their implications for innovation, with broad relevance for research policy and talent development.

1. Introduction

Scientific research is a dynamic and evolving endeavor driven by the pursuit of knowledge and innovation (Azoulay et al., 2019, 2018; Fortunato et al., 2018; Liu et al., 2023). A key aspect of this dynamism is scientists' tendency to explore new research directions, often through topic-switching (Huang et al., 2022; Zeng et al., 2019). Topic-switching, the process of transitioning across different research areas over a career, plays a crucial role in advancing knowledge and generating groundbreaking ideas (Hill et al., 2021; Jia et al., 2017; Liu et al., 2021). Higher levels of topic switching are indicative of exploration, whereas lower levels reflect exploitation. Together, these two processes shape the dynamic trajectories of scientific careers and may be integral to driving innovation (Manso et al., 2023).

Previous studies have shown that research topic patterns can predict scientific impact, including research prominence (Huo et al., 2022), recognition through awards (Ma & Uzzi, 2018), grant success (Guo et al., 2024), and career advancement (Liu et al., 2018, 2021). While prior research has examined topic-switching across disciplines and its link to academic productivity (Huang et al., 2022; Liu et al., 2021; Zeng et al., 2019), there is limited understanding of its impact on scientific disruption (Funk & Owen-Smith, 2017; Park et al., 2023; Wu et al., 2019), novelty (Shi & Evans, 2023; Uzzi et al., 2013), and interdisciplinarity (Gates et al., 2019; Stirling, 2007).

This study aims to bridge these gaps by investigating the relationship between topic-switching and scientific innovation.

* Corresponding author. School of Information Management, Nanjing University, Nanjing, 210023, PR China
E-mail address: alexjieyang@outlook.com.

https://doi.org/10.1016/j.ipm.2025.104171
Received 3 March 2024; Received in revised form 8 February 2025; Accepted 24 March 2025
Available online 1 April 2025
0306-4573/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Specifically, it explores how topic-switching influences the disruptive potential, novelty, and interdisciplinary nature of research. I hypothesize that scientists who switch topics more frequently are more likely to drive disruptive advancements and produce novel ideas that transcend disciplinary boundaries.

This study addresses the following key questions:

(RQ1) How have topic-switching trends evolved over time, and how do they vary across scientific fields?

(RQ2) How do topic-switching behaviors differ among scientists at various career stages, levels of prominence, and between male and female researchers?

(RQ3) What is the relationship between topic-switching and scientific innovation, including disruption, novelty, and interdisciplinarity?

To answer these questions, I analyzed a dataset of 27.6 million research papers authored by 1.4 million scientists from 1950 to 2020. To ensure data accuracy, I focused on first and last authors—where the last author typically serves as the corresponding author (Shi et al., 2023; Zhou et al., 2024)—and included only scientists with at least 10 first- or last-authored papers. Using advanced methodologies, I quantified scientific innovation through measures of disruption, novelty, and interdisciplinarity.

This research offers theoretical and practical contributions by revealing significant differences in topic-switching tendencies across disciplines, career stages, and demographic groups. Notably, elite scientists switch topics less frequently than their peers, and female scientists switch slightly more often than males. Findings suggest that while topic-switching promotes innovation, it may also hinder career success due to lower citation counts. These insights inform strategies for fostering a research culture that balances exploration, innovation, and career stability.

The remainder of this paper is structured as follows: Section 2 reviews relevant literature on topic selection and innovation. Section 3 describes the dataset and methodology. Section 4 defines key metrics, including switch scores and innovation indicators. Section 5 presents trends in topic-switching behavior. Section 6 explores the link between topic-switching and innovation. Section 7 discusses theoretical and practical implications. Finally, Section 8 concludes the paper, highlighting its limitations and potential directions for future research.

2. Theoretical background

2.1. Topic switching behavior of scientists

The selection of research topics significantly influences both individual careers and the broader scientific landscape. A key driver of scientific inquiry is the evolving interests of researchers, reflected in shifts across different domains (Fortunato et al., 2018). Sociologists of science have extensively studied the factors guiding topic selection, emphasizing the balance between conservative research and innovative risk-taking. They have also developed models categorizing research strategies to better understand topic selection dynamics (Foster et al., 2015; Shi et al., 2015).

The exploration-exploitation model explains scientists' research topic selection (Huang et al., 2022; Zeng et al., 2019). Exploration involves venturing beyond established expertise, which carries higher risk but offers opportunities for groundbreaking discoveries through novel combinations of ideas (Uzzi et al., 2013). Conversely, exploitation focuses on building upon existing knowledge, leading to consistent publications but potentially limiting originality (Foster et al., 2015; Lin et al., 2023; Shi et al., 2015).

Hill et al. (2021) introduced a framework to measure how far researchers move from their prior work when producing new studies. Their analysis of millions of scientific publications and patents reveals a "pivot penalty," where the impact of research declines as scientists deviate further from their previous topics. Zeng et al. (2019) examined topic-switching behaviors through co-citation network analysis, showing an increasing tendency for scientists to transition between topics over time, particularly as their careers progress. Liu et al. (2021) applied network analysis and deep learning to study career trajectories across art, filmmaking, and science. They found that individuals often explore diverse topics before experiencing a "hot streak" of high productivity, after which they focus on a narrower research area (Liu et al., 2018). This suggests that topic-switching tends to decrease as scientists advance in their careers.

Chen and Ding (2023) used the BERT model to vectorize Nobel laureates' papers, clustering research topics via Affinity Propagation. Guo et al. (2024) analyzed career trajectories of junior scientists in China, revealing that those who pursued new research directions in their initial funding cycle faced lower chances of securing future funding. Additionally, Venturini et al. (2024) investigated the link between topic-switching and collaboration, finding that inactive researchers are more likely to switch topics when working with a greater number of active coauthors.

Numerous studies have examined factors influencing topic-switching within the science of science field. Jia et al. (2017) developed the "seashore walk" model, using a random-walk approach to analyze topic selection behavior. Their findings indicate that research topic transitions follow an exponential distribution. Huang et al. (2023) expanded on this framework, demonstrating that shifting research interests correlate with productivity and impact.

While prior research has extensively investigated the relationship between topic switching, productivity, and scientific impact, its connection to more nuanced mechanisms of scientific innovation remains underexplored. Specifically, there is limited understanding of how topic switching influences multiple dimensions of innovation, including disruption, novelty, and interdisciplinarity—a gap this study aims to address.

2.2. Scientific innovation

Scientific innovation—the process of generating new ideas, discoveries, or technologies—drives progress across fields and shapes

modern society (Fortunato et al., 2018; Liu et al., 2024; Yang, 2024, 2024; Yang et al., 2025). It fuels economic growth, technological advancements, and social development (Yin et al., 2022; Zhang et al., 2024). Breakthroughs in one area often inspire progress in others, a phenomenon known as knowledge spillover (Ahmadpoor & Jones, 2017). Governments, funding agencies, and academic institutions play key roles in supporting innovation by encouraging risk-taking, collaboration, and knowledge sharing (Wei et al., 2023). Diversity in the scientific community further strengthens innovation by bringing in varied perspectives and expertise (Lu et al., 2022; Xu et al., 2024).

Researchers define scientific innovation in various ways, often emphasizing novelty, impact, and societal relevance (Acemoglu et al., 2018; Azoulay et al., 2018, 2011; Park et al., 2023; Uzzi et al., 2013). This builds on Schumpeter's idea that innovation involves introducing new products, processes, or organizational methods. Measuring scientific innovation is challenging due to its complexity. Traditional citation-based indicators—such as citation counts, the h-index, and field-normalized metrics—help assess research impact but often fail to capture true novelty or disruptive potential (Min et al., 2021; Savov et al., 2020; Alex Jie Yang, Linwei Wu et al., 2023). To address this, researchers have developed alternative methods. For example, Funk and Owen-Smith (2017) introduced the disruption score, which measures how much a study challenges existing knowledge. Uzzi et al. (2013) proposed the novelty score, which identifies unconventional combinations of prior knowledge. Stirling (2007) created the interdisciplinary diversity index, which assesses the degree of knowledge integration across different fields.

A key question in science policy is how creativity, innovation, and topic-switching behavior interact. Studies suggest that major breakthroughs often happen early in a scientist's career, largely due to high productivity at that stage (Liu et al., 2018; Sinatra et al., 2016). A model of scientific impact suggests that breakthroughs result from both skill and the ability to choose high-potential problems. Career progression, such as promotions and grants, often depends on consistent publishing, which favors incremental progress over risky new directions (Lim et al., 2025). While changing research topics can reduce productivity and pose career risks, sticking too closely to a single area may limit originality (Hill et al., 2021).Thus, exploring new topics can create fresh opportunities for innovation, suggesting that topic-switching—despite its risks—may lead to more groundbreaking discoveries.

3. Data

3.1. Scientific publication data

This study uses data from the Microsoft Academic Graph (MAG) (Wang et al., 2020), a comprehensive database of scholarly research. The dataset I analyzed is the version released on December 6th, 2021, which includes over 200 million documents published between 1800 and 2021, covering journal articles, conference proceedings, preprints, and other research outputs.

To focus on topic-switching behavior and its global trends, I limited the analysis to papers published from 1950 to 2020. This period saw major changes in the scientific landscape and aligns with previous studies (Lin et al., 2023). I further narrowed the dataset to papers with at least five references, resulting in a total of 45.2 million papers. This filtering served two purposes: first, to ensure that the papers contained enough reference information to calculate the "switch score" (Hill et al., 2021), and second, to exclude non-research articles and incomplete records. The final dataset included 27,556,282 research papers.

The MAG taxonomy categorizes research fields into two levels: the first level contains 19 broad categories (such as Medicine, Physics, and Computer Science), and the second level includes 292 subfields. Each paper in the dataset is assigned one or more labels, determined by machine learning models. These labels come with probability values, reflecting the model's confidence in its predictions. For the analysis, I used the second-level label with the highest confidence, especially when multiple labels were assigned. This classification system allows me to place each scientist into a specific field and identify interdisciplinary research teams.

3.2. Disambiguated author career data

MAG uses advanced methods like machine learning and crowdsourcing to address the issue of name disambiguation (Wang et al., 2020). It also gathers publicly available information from sources like personal websites and CVs to improve accuracy. A study by Lin et al. (2023) showed that MAG's disambiguation methods achieve 100 % accuracy and 84 % recall, demonstrating their effectiveness. This dataset provides detailed career information for scientists across all fields over the past 70 years.

To avoid the influence of incorrect authorship, I focused on the first and last authors, with the last author typically being the corresponding author. My analysis considered only scientists who had authored at least 10 papers as the first or corresponding author. To ensure data reliability, I excluded scientists with >500 publications to minimize name disambiguation errors (Yang & Wang, 2024). As a result, I obtained comprehensive career data for 1460,121 unique scientists.

I determined each scientist's "home field" based on the most common research areas of their first- and last-authored papers, using both first-level and second-level categories from the MAG taxonomy. Each scientist was assigned a single home field.

To infer gender, I used a statistical model by Van Buskirk et al. (2023), which associates first names with gender probabilities. This method has an accuracy range of 72 %−76 % (Van Buskirk et al., 2023). Gender probabilities were assigned to 1290,442 out of 1436, 437 scientists in the dataset. A probability above 0.5 indicates a female author, while lower probabilities indicate male authors. The final dataset includes 7246,749 papers written by 337,538 female scientists and 25,566,791 papers written by 952,904 male scientists. It's important to note that the gender inference algorithm may perform poorly for certain groups, such as those from Asia. However, since gender is not a primary focus of this study, I did not incorporate it into the regression models. As a result, any conclusions drawn from this gender analysis should be interpreted with caution.

4. Methodology

4.1. Quantifying topic switching

The goal of this study is to measure how much scientists shift their research topics throughout their careers. Rather than simply tracking whether a scientist changes topics, I focused on quantifying the degree of change for each paper they author (as the first or last author). I treated each journal as a sub-topic, as journals generally represent specific research areas. While interdisciplinary journals may cover multiple fields, they make up a small portion of the data. In total, the dataset includes 39,893 distinct journals, and each paper is represented by a knowledge distribution vector of length 39,893. This approach provides a comprehensive view of the knowledge base across all 27,556,282 papers in the dataset and has been used in previous research (Hill et al., 2021; Peng et al., 2021; Shi & Evans, 2023; Uzzi et al., 2013). This method allows for the analysis of topic shifts for scientists over their entire careers, spanning the past 70 years.

Following the study of Hill et al. (2021), I utilized a cosine-similarity metric to measure the extent to which a given paper departs from a scientist's prior body of work. Specifically, as shown in Fig.1, for a scientist and a focal paper *fp*, I calculated a vector $ks_{fp} = \left\{j_{fp}^{1}, j_{fp}^{2}, \ldots j_{fp}^{39893}\right\}$, representing the knowledge distribution by *fp* (i.e., the distribution of journals referenced by *fp*). Similarly, I counted the frequency in which different knowledge distributions appear in the union of prior work of the scientist in the past three years, defining a vector $ks_{past3} = \left\{j_{past3}^{1}, j_{past3}^{2}, \ldots j_{past3}^{39893}\right\}$. The switch score is then defined as:

$$Switch\ score = 1 - CosSim\left(ks_{fp}, ks_{past3}\right) \tag{1}$$

where $CosSim\left(ks_{fp}, ks_{past3}\right)$ is the cosine similarity between vector ks_{fp} and vector ks_{past3}. The switch score ranges from 0 (indicating no switch at all) to 1 (indicating maximal switch), adequately considering and controlling for the interdisciplinary nature of research papers. A switch score of 0 implies that *fp* draws on the exact same distribution of knowledge as the author's prior work, while a switch score of 1 indicates that *fp* draws entirely on a novel set of knowledge. This methodology enables me to capture the extent of departure in research focus while accommodating the interdisciplinary nature of scholarly output. By computing switch scores for each paper, I constructed a comprehensive profile of scientists' career trajectories, facilitating a nuanced analysis of topic shift patterns across diverse research domains.

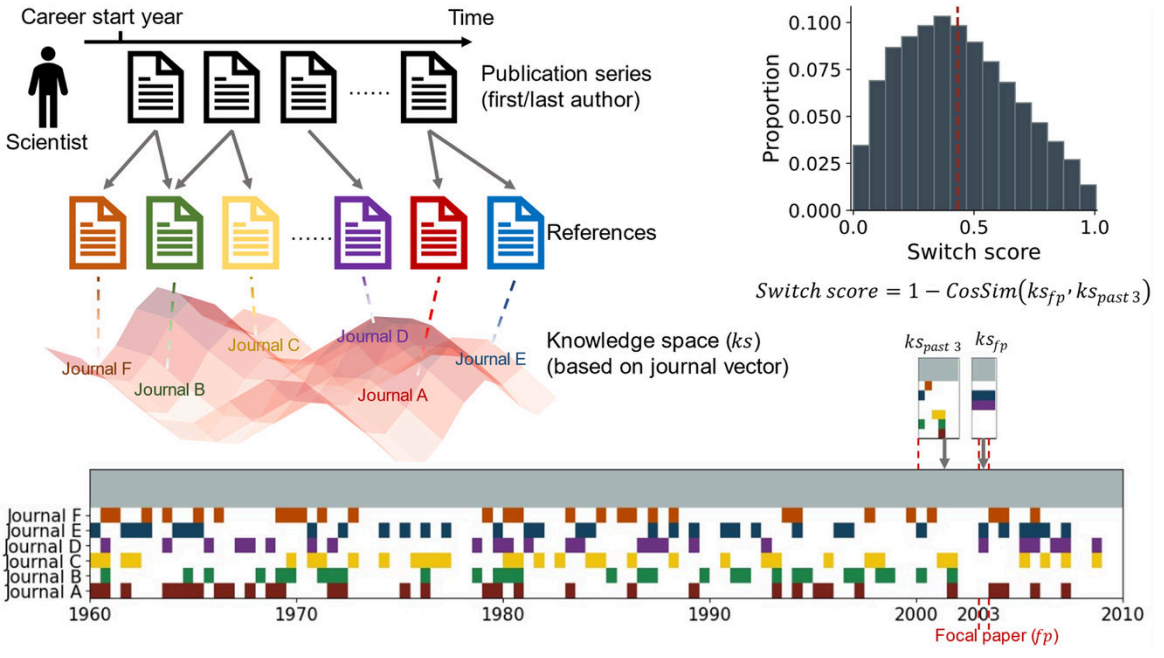


Fig.1. Visualization depicting the knowledge space trajectory of a scientist's career and the computation of the switch score. Each publication contributes to the expansion of the knowledge space, represented as a vector derived from the journal references within the paper. The switch score, computed as the cosine distance, assesses the alignment between the knowledge space of a focal paper and that of all papers published in the preceding three years of a scientist.

4.2. Null model: reshuffled data

The switch score proposed in this study has intrinsic limitations, as it is mainly based on the distribution of journals in the cited papers to denote the knowledge space trajectory of a scientist's career. First, referencing behavior changes over time and varies across different years. Second, journals exhibit different patterns; while some fields may have journals with a broader scope that produce a large number of papers annually, other fields may have more specialized journals that produce fewer papers. The size and number of journals forming the core of each field will vary across time and fields. Although MAG is sufficiently large to capture the switch score patterns of different types of scientists across various years and fields, direct comparisons of the switch score in different contexts (e.g., different years, different fields) are problematic. This complicates the identification of overall patterns in scientists' topic-switching behavior. To address this issue, I employed a null model in this study.

In the context of the science of science, a null model is used to establish a baseline for assessing the significance of observed patterns by reshuffling the data (Cimini et al., 2019). The null model method is widely applied in network-based or career level analyses (Liu et al., 2018; Park et al., 2023; Zeng et al., 2019). This process involves randomizing key attributes while preserving the overall distribution and structural properties of the original dataset, thereby dismantling any inherent correlations. Such a model serves as a crucial benchmark, allowing researchers to distinguish between meaningful scientific phenomena and those that could arise by chance. By comparing empirical results against outcomes derived from the null model, scholars can rigorously evaluate the robustness of their findings, ensuring that the detected relationships reflect genuine underlying mechanisms within the scientific landscape rather than random variation.

Specifically, I randomly reshuffled the paper-journal data while controlling for the publication year; that is, I switched the journal of a paper in one year to a different, randomly selected paper. Based on the reshuffled data, I constructed new author career datasets and recalculated their switch scores for each paper as defined in Section 4.1. This new switch score based on the reshuffled data represents the null model.

4.3. Scientific innovation measures

4.3.1. Scientific disruption

The disruption score, a recently proposed yet extensively verified measure, assesses the extent to which an idea disrupts the state of science or technology (Funk & Owen-Smith, 2017; Lin et al., 2023; Park et al., 2023; Wu et al., 2019). As illustrated in Supplementary Fig. S1, the disruption score is defined by the equation:

$$Disruption\ score = p_D - p_C = \frac{n_i - n_j}{n_i + n_j + n_k} \tag{2}$$

where n_i (the number of disruptive citations) represents the count of publications citing the focal paper without referencing its cited sources, n_j (the number of consolidating citations) denotes the count of publications citing the focal paper and its references, and n_k represents the count of publications citing the focal paper 's references without citing the focal paper itself.

A lower disruption value indicates extensive citation of references by papers referencing the focal paper, aligning with established knowledge and consolidating existing paradigms. Conversely, higher disruption values signify highly disruptive papers capable of inducing paradigm shifts. Understanding disruption sheds light on knowledge dissemination patterns and scientific paradigm shifts (Alex Jie Yang, Sanhong Deng et al., 2023, 2023). To address the centralized distribution of disruption scores around 0 (Yang, Gong et al., 2024), I also employed a binary variable where disruption > 0 as the alternative measure, which also proves successful in recent studies (Lin et al., 2023).

4.3.2. Scientific novelty

To quantify the novelty of papers, I examined the manifestation of atypical combinations of knowledge (Uzzi et al., 2013). This metric is operationalized by scrutinizing the distribution of journal pairs within a paper's reference list, as shown in Supplementary Fig. S2. Each journal pairing is transformed into z-scores, representing standardized representations. Novelty, indicating atypical combinations of knowledge, is quantified through Monte Carlo simulations. These simulations involve creating reshuffled networks with random edge reassignment while preserving the temporal and distributional attributes of the original citation network. The computation of atypical combinations for each journal pair is encapsulated by the formula:

$$Z\ score_{i,j} = \frac{obs(pair_{ij}) - exp(pair_{ij})}{\sigma(pair_{ij})} \tag{3}$$

where $obs(pair_{ij})$ signifies the observed frequency of the journal pair in the real dataset, $exp(pair_{ij})$ represents the mean, and $\sigma(pair_{ij})$ denotes the standard deviation of journal pairs obtained from 10 randomized simulations of the reshuffled network. To encapsulate the information within the distribution, denoted as a set $\{Z\ score_{i,j} \mid i.j \in J\}$, wherein J encompasses all journals within the reference list, I leveraged the 10th percentile value of the z scores set as a succinct summary statistic.

$$Novelty = \begin{cases} 1, if\ 10pct\{Z\ score_{i,j} \mid i.j \in J\} < 0 \\ 0, if\ 10pct\{Z\ score_{i,j} \mid i.j \in J\} \geq 0 \end{cases} \tag{4}$$

In this study, a paper is classified as novel if its z-score, representing its rank within the 10th percentile of the distribution, is <0. Papers that do not meet this criterion are not categorized as novel.

4.3.3. Interdisciplinarity

Interdisciplinarity is a crucial facet of contemporary research, reflecting the dynamic fusion of diverse knowledge sources within scholarly discourse. To measure this multifaceted quality, I adopt the Rao-Stirling index (Gates et al., 2019; Stirling, 2007). This index serves as a robust measure, encapsulating the intricate interplay between disciplines within scholarly works.

For each paper i and its reference list $R_i = \{r1, r2, \ldots\}$, I commenced by mapping its references into a vector $F_i = \{field_{r1}, field_{r2}, \ldots, field_{rk}, \ldots\}$, where the $field_{rk}$ signifies the field of reference $rk \in R_i$, delineated at the second level within the MAG. Subsequently, I focused on a specific field m and consider all papers published therein during year t. I aggregated the paper vectors within this field, and based on its citing relationship with the other 291 fields, to construct the knowledge vector V_m for field m. Utilizing the cosine distance metric, I then computed the distance between fields m and n as $d_{mn} = 1 - CosSim(V_m, V_n)$.

The Rao-Stirling index, fundamental to my analysis, is defined by the following formula:

$$Rao - Stirling = \sum_{m \neq n} d_{mnt} p_m p_n \tag{5}$$

where p_m denotes the fraction of references in discipline m, and d_{mnt} is the distance between field m and n in year t. This index serves as a comprehensive measure, reflecting the diversity of knowledge sources inherent in research endeavors. With 292 second-level disciplines cataloged within MAG, my metric adeptly captures the nuanced balance, variety, and disparity present within scholarly works. Scores ranging from 0 to 1 signify varying degrees of diversity, delineating the interdisciplinary nature of research papers with precision and nuance.

4.4. Control variables

To account for factors that might influence topic-switching patterns and scientific innovation, I included several control variables at the paper, team, and author levels. These controls help isolate the specific impact of topic switching on innovation, providing a more accurate analysis.

At the paper level, I considered the influence of funding from key organizations such as the National Institutes of Health (NIH) and the National Science Foundation (NSF). These funding sources could impact the direction and outcomes of scientific research, and thus, their inclusion as a control allows for a better understanding of their potential role in research productivity (Yang, 2024; Zhai et al., 2024). Additionally, I controlled for the number of references in each paper, as this can reflect the breadth of knowledge a paper engages with and may influence both the switch score and innovation metrics.

To address potential differences across scientific fields, journals, and over time, I incorporated fixed effects for scientific domains, years, and journals. The dataset's 292 second-level fields were used to account for variations across different research areas, recognizing that scientific innovation and research dynamics can differ significantly by discipline (Chu & Evans, 2021). I also included year fixed effects to account for temporal trends, ensuring that shifts in scientific innovation are captured across different time periods (Park et al., 2023). Lastly, journal fixed effects were used to control for the inherent heterogeneity of publication venues, as the diversity of journals can influence research outcomes. The total of 39,893 journals in the dataset ensured a comprehensive control for this factor.

At the team level, I took into account the collaborative nature of modern science. Over the years, large, interdisciplinary teams have become essential to tackle complex research questions, and I controlled for the number of authors involved in a paper to reflect the extent of collaboration. Recognizing the global nature of research today, I also included a variable for international collaboration, considering the authors' country affiliations. Another important factor is the diversity of knowledge within a research team. I measured the multi-disciplinarity of teams by identifying the primary research field of each team member, distinguishing between monodisciplinary teams (where all members come from the same field) and multidisciplinary teams (where team members represent diverse fields). This allowed me to capture the breadth of knowledge available to the team and its potential impact on the research outcomes (Lin et al., 2023).

At the author level, I used several measures to assess the experience and academic standing of the scientists. Author academic age was calculated based on the number of years since a scientist's first publication, ranging from 0 to 60 years. This variable helps to account for how career stage influences research decisions and topic switching (Xu et al., 2024; Yang, Xu et al., 2024). Author productivity was measured by the total number of papers authored up to the focal paper's publication date, while author citations and the h-index were used to gauge an author's influence and research impact (Alex Jie Yang, Linwei Wu et al., 2023). These metrics provide insights into the researcher's track record, which may affect their approach to innovation and topic-switching behavior.

To address potential selection biases in my analysis, I included author-level fixed effects in the regression models. This was done by creating a dataset of paper-author observations, totaling 35,786,731 entries, with characteristics of both first and last authors. Including these fixed effects ensures that the analysis accounts for individual author characteristics, helping to isolate the effect of topic switching on innovation and providing a more accurate evaluation of its impact.

4.5. Empirical specification

To analyze the intricate relationship between topic switch patterns and scientific innovation, I employed multivariate fixed-effect

Ordinary Least Squares (OLS) regression models. Eqs. (6) estimate disruption based on these models as follows:

$$Disruption_i = \alpha + \beta_1(Switch\ score_i) + \sum_{k=2}^{K}\beta_k\left(Control_{k,i}\right) + \gamma_t Y_t + \delta_f F_f + \theta_j J_j + \alpha_a A_a + \varepsilon_i \tag{6}$$

where i represents the article, t the year of publication, f the field, j the journal, and a the authors. All counting variables are natural log-transformed within the models. The model structure accounts for numerous factors: Y_t denotes the year fixed effects, F_f captures second-level field fixed effects, J_j represents journal fixed effects, A_a signifies author fixed effects, and ε_i denotes the error term.

To address the concentration of disruption scores around zero, I also adopt the binary variable disruption>0 as an alternative measure, indicating that 19.5 % of papers are disruptive. In this regard, I utilized a fixed-effect logistic regression model to assess the impact of the switch score on disruption:

$$P(Disruption_i > 0) = \frac{1}{1 + e^{-\left(\alpha+\beta_1(Switch\ score_i)+\sum_{k=2}^{K}\beta_k\left(Control_{k,i}\right)+\gamma_t Y_t+\delta_f F_f+\theta_j J_j+\alpha_a A_a\right)}} + \varepsilon_i \tag{7}$$

where i represents the article, t the year of publication, f the field, j the journal, and a the authors.

I employed similar models to examine the impact of topic switching on interdisciplinarity and novelty in scientific papers. For interdisciplinarity, I utilized the fixed-effect Ordinary Least Squares (OLS) regression model to estimate the effect of switch score on Rao-Stirling:

$$Rao - Stirling_i = \alpha + \beta_1(Switch\ score_i) + \sum_{k=2}^{K}\beta_k\left(Control_{k,i}\right) + \gamma_t Y_t + \delta_f F_f + \theta_j J_j + \alpha_a A_a + \varepsilon_i \tag{8}$$

For novelty, I employed a fixed-effect logistic regression model to assess the impact of the switch score on novelty:

$$P(Novelty_i = 1) = \frac{1}{1 + e^{-\left(\alpha+\beta_1(Switch\ score_i)+\sum_{k=2}^{K}\beta_k\left(Control_{k,i}\right)+\gamma_t Y_t+\delta_f F_f+\theta_j J_j+\alpha_a A_a\right)}} + \varepsilon_i \tag{9}$$

5. Dynamic patterns of switch

5.1. Topic switching patterns of scientists

I started by examining the topic-switching behaviors of a randomly selected group of 20 scientists, each with at least 50 first- or last-

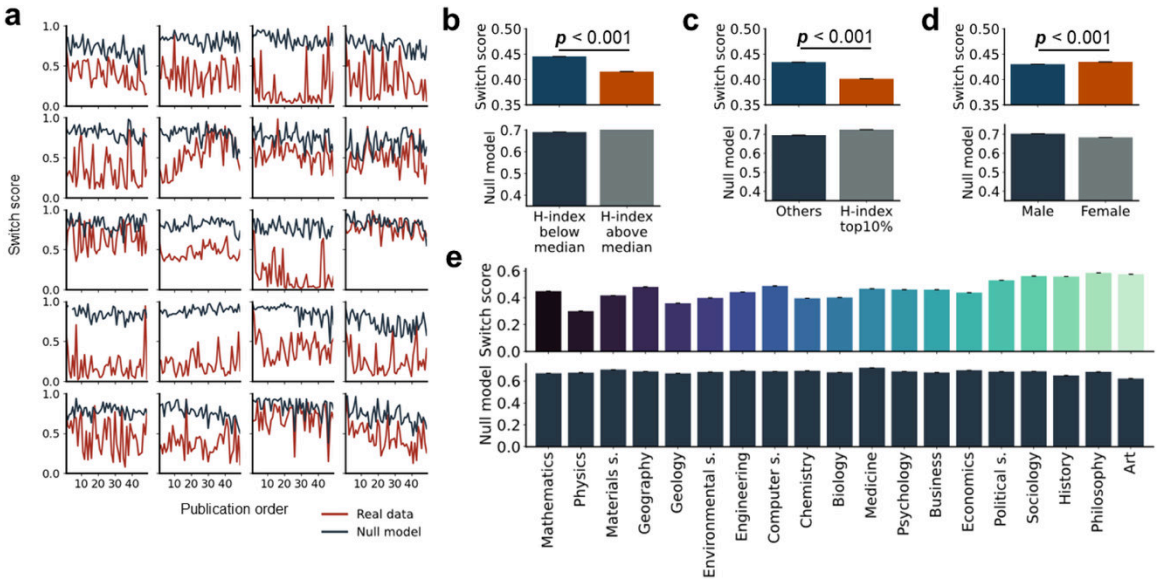


Fig. 2. Switch patterns of scientists. (a) Evolution of switch scores across careers for a randomly selected sample of 20 scientists, highlighting the patterns and variability in topic switch patterns. (b) Average switch scores and null models categorized by h-index below or above the median, indicating lower switch scores among scientists high h-index. (c) Average switch scores and null models categorized by h-index below or above the top 10 %, indicating lower switch scores among scientists in the top 10 % h-index bracket. (d) Average switch scores and null models of male and female scientists, with a slight difference indicating slightly higher switch scores among female scientists. (e) Average switch scores and null models across different fields of study. Error bars represent 95 % confidence intervals, and two-tailed Welch's *t*-tests are utilized to compare switch scores in all categories.

authored publications, as shown in Fig. 2a. The red lines represent the actual data, while the black lines represent a null model for comparison. For most scientists, their switch scores were consistently lower than expected, suggesting that they changed topics less frequently than the null model predicted. However, the scientists displayed a range of behaviors: some maintained high switch scores, others had consistently low switch scores, and a few showed significant fluctuations over time.

Next, I looked at whether a scientist's academic status, measured by their dynamic h-index, influenced their tendency to switch topics. Scientists were grouped based on whether their h-index was above or below the median in their field and year. The results, shown in Fig. 2b, revealed a surprising pattern: scientists with higher h-indices, indicating greater prominence, were less likely to switch topics. The null model, however, showed no significant difference in switch scores between these groups. I also categorized scientists who ranked in the top 10 % of their field's h-index. As shown in Fig. 2c, these elite scientists had lower switch scores compared to others, further confirming this trend.

When I compared topic-switching behavior by gender, I found small but notable differences. As shown in Fig. 2d, female scientists had slightly higher switch scores than male scientists. While this difference was statistically significant and consistent with the null model, the effect size was modest, suggesting that gender is not a major factor in topic-switching behavior.

Finally, I examined how topic-switching behaviors varied across different research fields. Fig. 2e shows the average switch scores for 19 broad fields, based on the MAG field taxonomy. The null model indicated little variation, but real data revealed clear differences across fields. Disciplines like Physics and Chemistry showed lower rates of topic switching, while fields such as Art and History had higher rates of topic switching.

5.2. Temporal trends

To understand the overall trend in topic-switching behavior across all fields, I analyzed data from 27,556,282 research papers authored by 1460,121 scientists. By calculating the average switch scores for each year and comparing them with null models, I identified temporal patterns.

As shown in Fig. 3, real switch scores decline over time. The null models, however, show a slight increase before 2000, followed by a decline after that. This suggests that the observed decline in switch scores may be due to changes in reference patterns or the distribution of journals within the citation system. To highlight the true trend, I compared the real data with the null model scores. Although z-scores would be the ideal method, the computational complexity of my approach made this unfeasible. Nevertheless, the difference between the real data and the null model effectively reflects the trend. The results in Fig. 3d show a decline in topic switching before 2000, followed by an increase afterward.

To further explore the trend in switch scores, I examined topic-switching behavior across different fields. The real data results are shown in Fig. 4a, while the null model results are in Fig. 4e. These findings reveal that scientists in science and engineering fields are less likely to switch topics, while those in the arts and humanities tend to change topics more frequently. Fig. 4i, which shows the difference between the real data and the null model, highlights the true trend: topic-switching in social sciences and science and engineering declined before 2000 but increased afterward. In contrast, the arts and humanities fields show a stable trend over time.

I also compared the topic-switching behavior of elite versus non-elite scientists based on their dynamic h-index. Elite scientists with a high h-index consistently exhibit lower levels of topic-switching than their non-elite counterparts. While their real data patterns differ (Fig. 4b and 4c), the difference between the real data and the null models (Fig. 4j and 4k) shows that both groups follow a similar trend: a decline in switching followed by an increase.

Similarly, male and female scientists exhibit the same topic-switching behavior (Fig. 4l), with the gap between genders narrowing over time. This indicates a convergence in topic-switching behavior between male and female scientists.

5.3. Career patterns

To analyze switch score trends at different career stages, I categorized scientists based on their career progression, starting from the year they published their first paper. The results from the real data, shown in Fig. 5a, indicate that early-career scientists tend to have

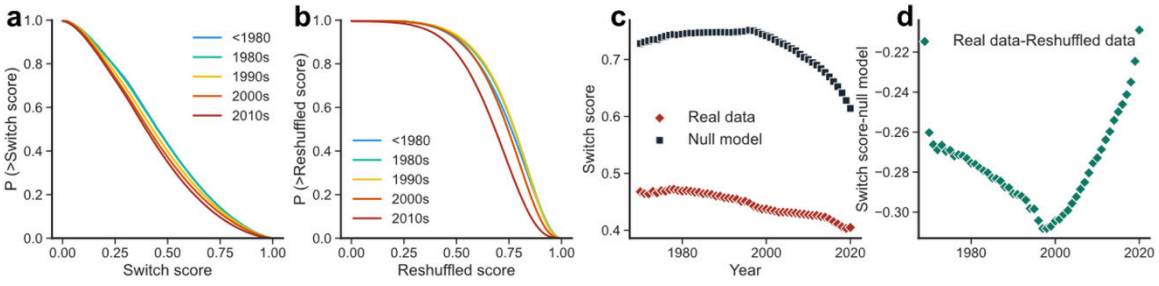


Fig. 3. Temporal evolution of the topic switch pattern in science. (a) The Complementary Cumulative Distribution Function (CCDF) of switch scores for each period demonstrates a decreasing likelihood of scientists switching their research topics over time in the real data. (b) The CCDF of randomly reshuffled switch scores for each period. (c) Both the real data and the null model of switch scores exhibit a decreasing trend over time. (d) The difference between the real data and the null model indicates a decrease in topic switching before 2000, followed by an increase after 2000.

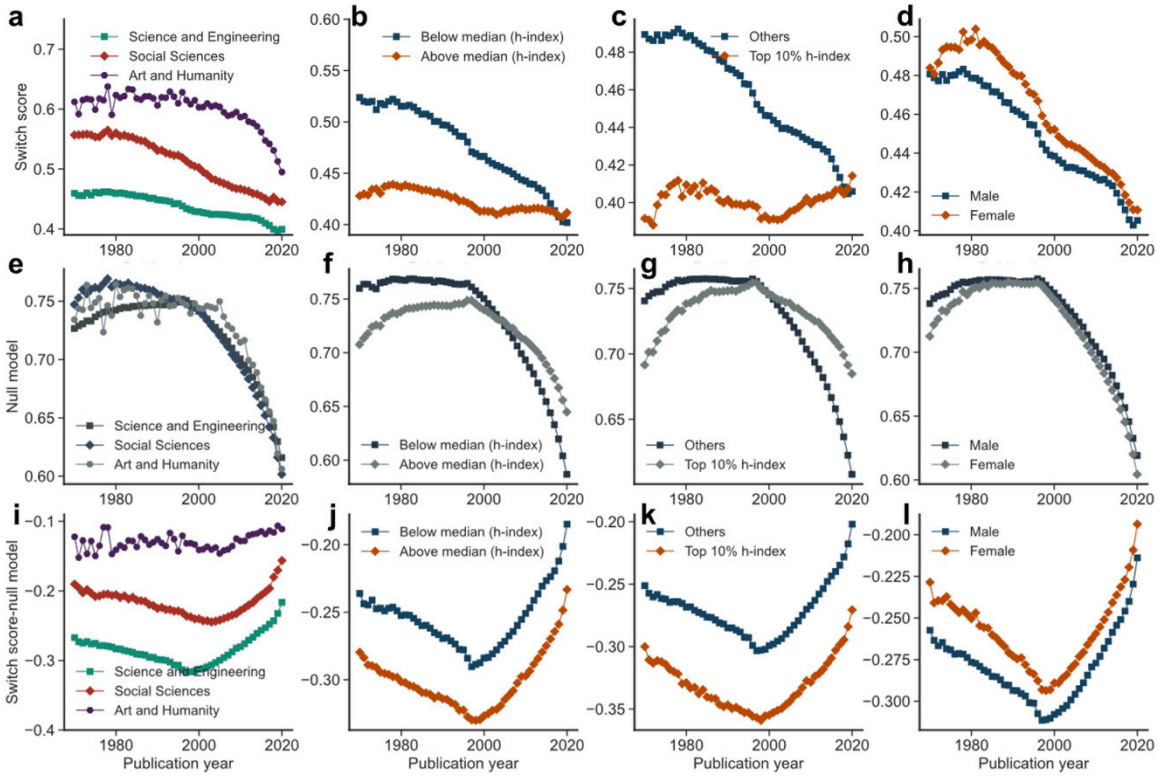


Fig. 4. Temporal evolution of topic switch pattern across various types of scientists. (a, e, i) Trends in switch scores across various fields of study. (b, f, j) Trends in switch scores for scientists with an h-index above and below the median. (c, g, k) Trends in switch scores for scientists in the top 10 % h-index bracket versus those who are not. (d, h, l) Trends in switch scores for male versus female scientists. (a-d) Average switch scores from the real data. (e-h) Switch scores from the null model with reshuffled data. (i-l) Differences between the real data and the null model.

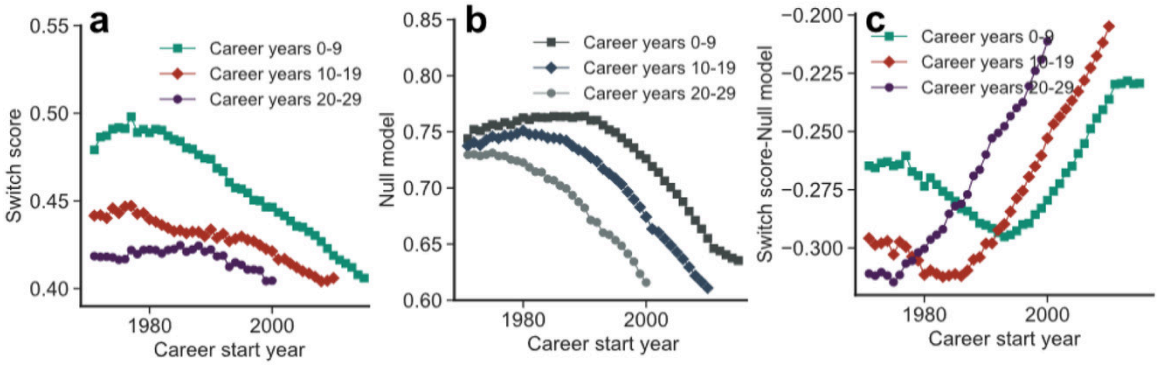


Fig. 5. Dynamics of topic switch pattern over the career start year of scientists. Switch scores with the career start years for scientists at different stages of their careers (the first 10 years, 10–19 years, 20–29 years). (a) Average switch scores from the real data. (b) Switch scores from the null model with reshuffled data. (c) Differences between the real data and the null model.

higher switch scores than those in the middle or late stages of their careers. However, as seen in Fig. 5c, this pattern is significant only before 1980. After that, middle- and late-career scientists show more consistent and faster increases in their topic-switching levels, while early-career scientists' topic-switching levels decrease until around 2000, then increase. Over time, early-career scientists show lower switch scores compared to those in the middle or late stages.

I also explored how the number of publications influences topic-switching behavior. Focusing on scientists with at least 100 publications, I started from their 10th paper to ensure accurate switch score calculation (see Fig. S5). The analysis shows that highly productive scientists tend to have more stable topic-switching patterns throughout their careers, although there is significant variation across fields. When comparing elite scientists (based on h-index) to their less distinguished peers, I found that early-career differences in topic-switching are minimal. However, as their careers progress, the gap widens, with elite scientists showing lower levels of topic

switching in their middle and late career stages.

6. Association between topic switching and scientific innovation

6.1. Descriptive statistics

I used an author-paper dataset that includes information on papers, coauthors, and authors, calculated up to the publication year of each paper. The dataset contains 35,786,731 observations, representing 1436,437 unique scientists and their 27,556,282 research publications from 1950 to 2020.

Table 1 presents the descriptive statistics of the variables under study. The switch score has a mean of 0.43 with a standard deviation of 0.23. On average, papers receive 36.21 citations and include 33.56 references, with a minimum reference count of 5. Approximately 10 % of the papers receive funding from either the NIH or NSF. The average team size is 4.59 members, and 19 % of the papers involve international collaboration. The disruption score ranges from −0.98 to 1, with a mean of 0 and a standard deviation of 0.01. Due to the centralized distribution of disruption scores around 0, I employed a binary variable where disruption > 0, indicating that 19.5 % of papers are considered disruptive. Additionally, I used a binary variable to indicate whether a paper represents a novel combination of knowledge based on its atypicality scores; approximately 0.9 % of papers are classified as novel (atypicality score < 0).

6.2. Relationship between topic switching and scientific disruption

To examine the relationship between topic switching and scientific disruption (Funk & Owen-Smith, 2017; Park et al., 2023; Alex Jie Yang, Sanhong Deng et al., 2023, 2023; Alex Jie Yang et al., 2023), I used both multivariate fixed-effect regression models and logit regression models. The results are shown in Fig. 6, with detailed regression effects and standard errors provided in Supplementary Tables S1 and S2.

I began by using a fixed-effect OLS regression model to estimate the impact of topic switching on scientific disruption. To account for potential confounding factors, I included controls at the paper, team, and author levels, as well as four-level fixed effects. My findings consistently showed a positive relationship between switch scores and disruption levels. Additionally, I found that factors like reference count, team size, international collaboration, productivity, and author citations were negatively associated with disruption, while multidisciplinary teams and career age were positively associated with disruption. Including these variables, along with fixed effects for year, field, author, and journal, improved the model's explanatory power (see Supplementary Table S1).

Given that disruption scores are mostly centered around 0, I also used a dummy variable to measure disruption (where disruption $>$ 0). Using fixed-effect logit regression models, I found that topic switching consistently has a positive effect on disruption. In the preferred model (Model 12), I found that an increase of 1 standard deviation in the switch score (0.23) made a paper about 39 % more likely to be disruptive (log ratio = 1.39). This suggests that scientists who frequently switch research topics are more likely to produce disruptive, innovative contributions in their fields.

6.3. Relationship between topic switching and novelty

I also used multivariate fixed-effect regression models to explore how topic switching influences interdisciplinarity and novelty in scientific research. The results are shown in Fig. 7, with specific regression coefficients and standard errors detailed in Supplementary Tables S3 and S4.

First, I applied a fixed-effect OLS regression model to assess how topic switching affects interdisciplinarity, measured by the Rao-Stirling index. To control for other variables, I included paper, team, and author-level controls, as well as four-level fixed effects. The results showed a positive relationship between topic-switching and interdisciplinarity. Specifically, an increase of 1 standard deviation in the switch score (0.23) was associated with a 0.05 increase in the Rao-Stirling score (Model 6).

Table 1
Descriptive statistics of the author-paper dataset.

Variables	Obs.	Mean	Std.	Min	Median	Max
Switch score	35,763,144	0.43	0.23	0.00	0.41	1.00
Disruption score	33,824,503	0.00	0.01	−0.98	0.00	1.00
Interdisciplinarity (Rao-Stirling)	35,581,027	0.21	0.08	0.00	0.22	0.43
Atypicality score	34,892,322	24.50	212.19	−445.27	−0.28	78,097.00
Citation count	35,786,731	36.21	138.10	0.00	14.00	218,903.00
Grant funding	35,786,731	0.10	0.30	0.00	0.00	1.00
Reference count	35,786,731	33.56	31.55	5.00	27.00	13,171.00
Team size	35,786,731	4.59	14.60	1.00	4.00	5123.00
International team (dummy)	35,786,731	0.19	0.39	0.00	0.00	1.00
Multidisciplinary team (dummy)	35,786,731	0.37	0.48	0.00	0.00	1.00
Author age	35,786,731	15.80	10.48	0.00	14.00	60.00
Author home field	35,786,731	0.72	0.45	0.00	1.00	1.00
Author past productivity	35,781,942	57.39	65.98	0.00	34.00	499.00
Author past citations	35,786,731	1078.34	2599.69	0.00	253.00	196,932.00

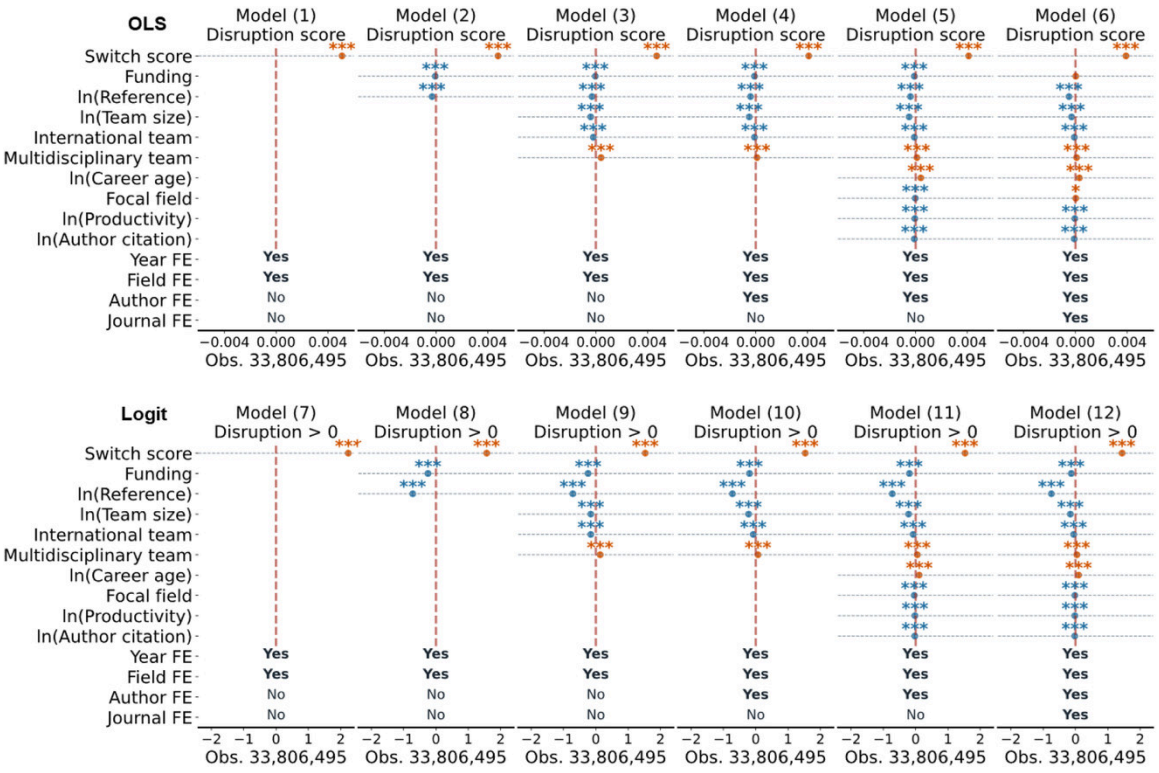


Fig. 6. Estimated effect of switch score on disruption. Models (1)-(6) utilize Ordinary Least Squares (OLS) regression models to estimate the effect of switch score on raw disruption scores. In all models, the coefficients of switch scores demonstrate significant positive effects. Models (7)-(12) employ logit regression models to estimate the effect of switch score on the probability of disruption being greater than 0. Once again, all coefficients of switch scores exhibit significant positive effects. For Models (1)-(3) and (7)-(9), robust standard errors are displayed as error bars. In Models (4)-(6) and (10–12), author fixed effects are included, and standard errors (error bars) are clustered at the author level. *p < 0.05; ** p-value < 0.01; *** p-value < 0.001.

Next, I used fixed-effect logit regression models to estimate how topic switching relates to scientific novelty. After including controls at the paper, team, and author levels, as well as four-level fixed effects, the analysis consistently showed a significant positive relationship between topic switching and novelty. In the preferred model (Model 12), an increase of 1 standard deviation in the switch score (0.23) made a paper 52 % more likely to be classified as novel (log ratio = 1.52). These findings indicate that scientists who frequently switch topics are more likely to engage in interdisciplinary research and use novel combinations of knowledge in their work.

6.4. *Heterogeneity and robustness*

To ensure the reliability of my findings, I conducted several sensitivity analyses. First, I divided the sample by the author's career age (ranging from 2 to 50 years) and evaluated the effect of the switch score on the likelihood of achieving a disruption score greater than zero, using the controls and fixed effects from Fig. 6 Model (12). The results, shown in Fig. 8a, demonstrate that the positive impact of topic switching on scientific disruption holds across all career stages.

Next, I assessed the stability of the results across different publication years, conducting a split-sample regression analysis covering the period from 1950 to 2020. Again, I used all relevant control variables and fixed effects. As shown in Fig. 8b, the positive relationship between topic switching and disruption remains consistent across different years. However, for papers from the 1950s, the results were insignificant due to limited publication data from that period.

I also explored the robustness of the findings across various academic fields by conducting a split-sample regression for different time periods within 19 broad fields. The results, shown in Fig. 8c, indicate that the positive effect of the switch score on disruption is consistent across fields. In the field of Art, however, the coefficients were not statistically significant, likely due to large standard errors. Nevertheless, the positive effect held across most fields.

While my research confirms the consistent positive link between topic switching and scientific disruption, there is evidence of a "penalty of pivot" in science (Hill et al., 2021), as shown in Fig. S7 and Table S5. This negative relationship suggests that switching topics may reduce citation impact. However, it's important to note that innovation is not solely measured by citations. Disruptive ideas often emerge from new perspectives and interdisciplinary work, which address complex scientific challenges. Thus, even if citation impact decreases when researchers shift their focus, their broader contribution to scientific progress can be more impactful than citation counts alone.

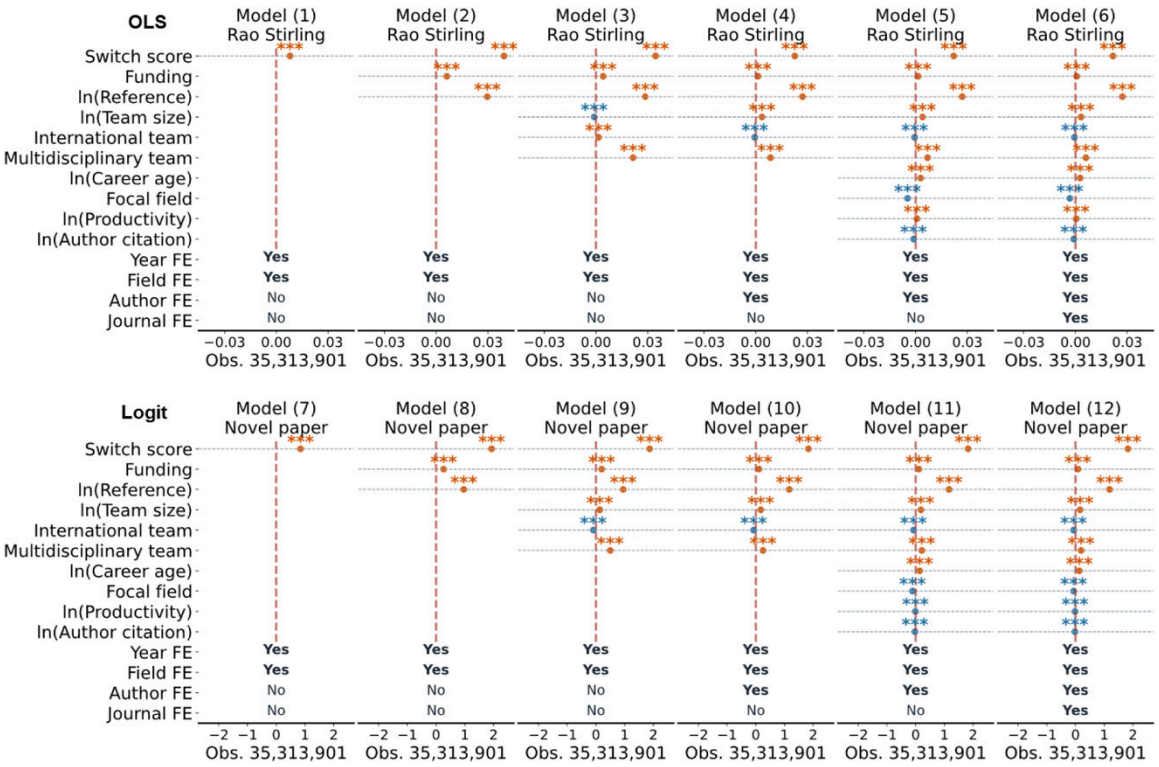


Fig. 7. Estimated effect of switch score on interdisciplinarity and novelty. Models (1)-(6) utilize Ordinary Least Squares (OLS) regression models to estimate the effect of switch score on the Rao-Stirling scores. All coefficients of switch scores are significantly positive except for Model (1). Models (7)-(12) employ logit regression models to estimate the effect of switch score on the probability of novel papers. All coefficients of switch scores are significantly positive. For Models (1)-(3) and (7)-(9), robust standard errors are depicted as error bars. For Models (4)-(6) and (10–12), author fixed effects are included, and standard errors (error bars) are clustered at the author level. *p $<$ 0.05; ** p-value $<$ 0.01; *** p-value $<$ 0.001.

7. Discussion

7.1. Theoretical implications

This study advances our theoretical understanding of scientific innovation by illuminating the critical role of topic-switching in shaping knowledge production. By demonstrating a significantly positive relationship between topic-switching and key indicators of scientific innovation—including disruption, novelty, and interdisciplinarity—these findings underscore the importance of intellectual exploration and boundary-crossing in fostering transformative research. Researchers who switch topics more frequently are more likely to challenge existing paradigms, introduce novel ideas, and integrate diverse perspectives, thereby accelerating the evolution of scientific knowledge. This highlights topic-switching as a vital mechanism for creative breakthroughs and the diversification of research landscapes.

The observed increase in topic-switching since 2000 suggests a shifting paradigm in scientific inquiry, where interdisciplinary approaches are gaining prominence. This shift may be driven by broader changes in research funding priorities, institutional incentives, and the increasing complexity of contemporary scientific challenges, all of which require collaboration across traditional disciplinary boundaries. As scientific problems become more interconnected—spanning fields such as artificial intelligence, climate science, and biomedical research—the ability to navigate across disciplines becomes increasingly valuable.

The finding that female scientists engage in topic-switching more frequently than their male counterparts offers important insights into gendered career dynamics in academia. This pattern may reflect differences in strategic career navigation, potentially shaped by disparities in access to resources, mentorship, and institutional stability. Women in science may adopt topic-switching as an adaptive response to systemic barriers, seeking new opportunities to establish their research impact in environments where traditional pathways to recognition and stability remain challenging. Similarly, the contrast between prominent and less prominent scientists in topic-switching behaviors suggests that intellectual mobility serves as both an opportunity and a risk. For early-career and less-established researchers, topic-switching may be a strategic means of carving out a unique research niche and adapting to shifting academic landscapes. However, the potential tradeoff between topic-switching and citation impact highlights a key tension in scientific careers—while intellectual flexibility fosters innovation, it may come at the cost of immediate visibility and recognition. This suggests that traditional academic success metrics, which prioritize citation-based influence within established fields, may undervalue the long-

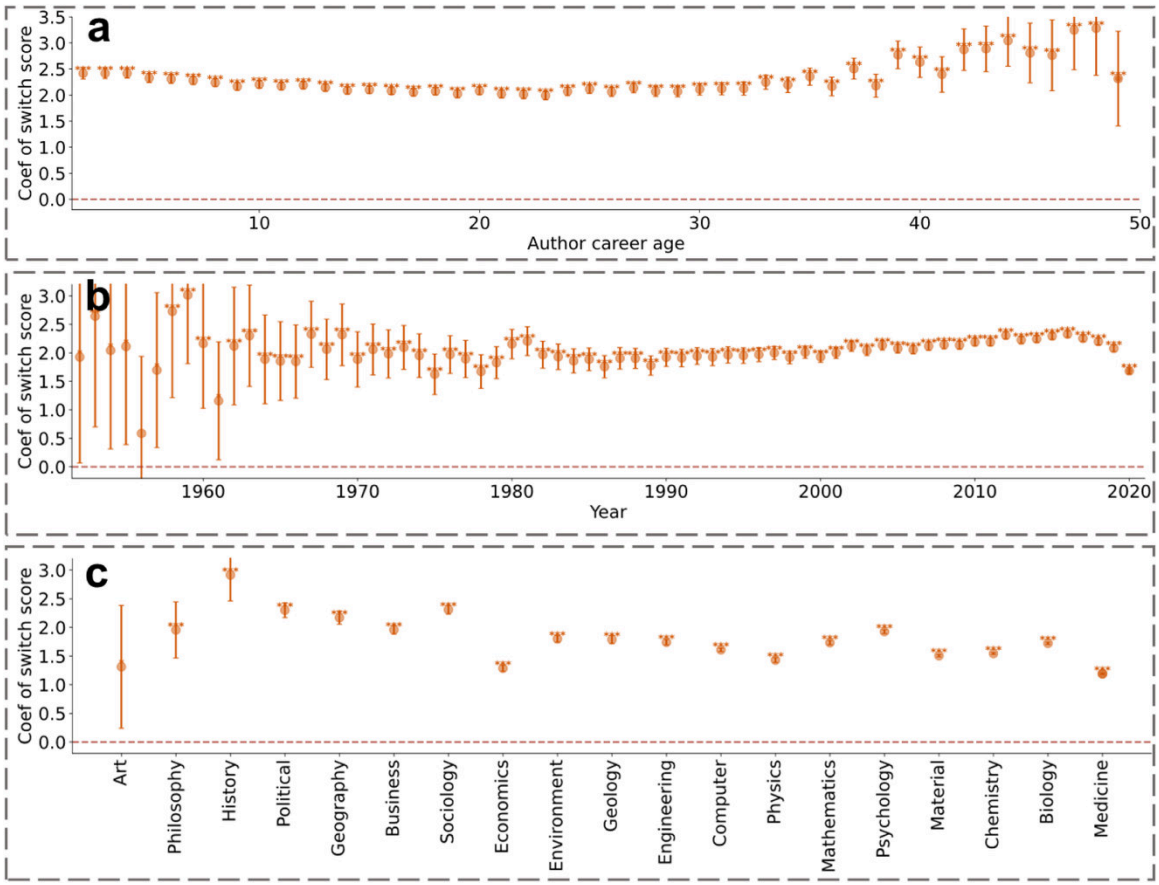


Fig. 8. Robust check for the effect of switch score on scientific disruption. (a) The positive impact of switch score on scientific disruption remains consistent across different career stages of authors. (b) The positive impact of switch score on scientific disruption is stable across different publication years of focal papers: results from yearly split-sample regression analysis. (c) The positive impact of switch score on scientific disruption is consistent across different fields of study: results from field-level split-sample regression analysis. Each node represents the logistic regression of switch score on the probability of having a disruption score greater than 0, with all confounding variables and fixed effects accounted for in each setting. The error bars depict the upper and lower bounds of the 95 % confidence intervals based on clustered standard errors at the author level. Statistical significance is denoted as *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

term contributions of researchers who pursue unconventional or interdisciplinary paths (Hill & Stein, 2025; Lim et al., 2025).

7.2. Practical implications

This study also highlights important practical implications for the scientific community. The findings suggest that switching research topics can lead to more innovative and disruptive ideas, but it also carries the risk of lower citation counts, potentially affecting researchers' success.

The dynamic nature of scientific inquiry calls for a research environment that encourages experimentation and exploration. Academic institutions play a key role in fostering interdisciplinary research and supporting unconventional research paths. For early-career researchers, it's important to nurture a mindset that values exploration, as this can lead to innovative breakthroughs. Evidence shows that scientists tend to explore a wide range of topics early in their careers, which often leads to scientific disruption. Therefore, academic institutions and mentorship programs should encourage young researchers to pursue new research directions by offering networking opportunities and promoting collaborative projects. This can enhance knowledge exchange and stimulate creativity among emerging scientists.

Encouraging topic switching could increase the novelty of research. Academic institutions can incentivize this by rewarding work that crosses traditional disciplinary boundaries. Policymakers and funding agencies also play a crucial role in supporting topic-switching research. Funding should prioritize projects that promote diversity and inclusion, as well as support underrepresented groups in science. Additionally, policymakers should support open science initiatives, advocate for data sharing, and promote collaboration to enhance knowledge dissemination across diverse research communities.

8. Conclusion

This paper explores how topic-switching behavior impacts scientific innovation, focusing on how changes in research direction influence contributions to disruption, novelty, and interdisciplinarity. The study finds that topic-switching has increased since 2000, with STEM fields showing lower levels of switching compared to the Arts and Humanities. It also uncovers gender differences, with female scientists switching topics slightly more often than male scientists, and a notable difference in switching behavior between prominent and non-prominent scientists. The analysis of career stages shows that early-career scientists are more likely to explore diverse topics, while more experienced scientists tend to specialize. Importantly, the study finds a positive relationship between topic-switching and scientific innovation, suggesting that shifting research focus leads to groundbreaking advancements, although it may lower citation impact.

While the study provides valuable insights into the patterns of topic-switching and its link to scientific innovation, several limitations should be considered. First, the analysis focused only on first-authored and last-authored papers, potentially overlooking the contributions of co-authors. Second, the MAG dataset has limitations, particularly in covering non-English publications and humanities fields. The process of author disambiguation may also introduce biases, affecting the results. Third, the study's observational nature means it cannot establish causality between topic-switching and innovation. The findings show associations, but alternative explanations may exist. Fourth, gender data may be inaccurate. Finally, the use of quantitative measures like disruption and novelty scores may introduce biases, as they may not fully capture the breadth of scientific innovation.

Despite these limitations, this study sets the stage for future research into the dynamics of scientific innovation. Future studies could use experimental designs to better understand the causal relationship between topic-switching and innovation. An interesting area for further research is the exploration of topic-switching behavior at different career stages. It would be valuable to examine whether senior researchers, who often rely on a core set of established readings, switch topics differently from junior researchers who are exploring new literature. A deeper analysis of career-stage differences could provide more insights into how topic-switching influences innovation and research impact. Studying how these patterns evolve over time could also shed light on how career progression shapes scientific productivity and novelty.

9. Declaration of competing interest

The author declares no competing interest relevant to this article.

CRediT authorship contribution statement

Alex J. Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Acknowledgments

This research was supported by the National Natural Science Foundation of China Young Student (Doctoral) Project (#724B2015). The study was conducted independently, and the author would like to express gratitude to anonymous reviewers for their valuable insights.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ipm.2025.104171.

Data availability

Data will be made available on request.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., & Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review, 108*(11), 3450–3491. https://doi.org/10.1257/aer.20130470

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science (New York, N.Y.), 357*(6351), 583–587. https://doi.org/10.1126/science.aam9527

Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2019). Does science advance one funeral at a time? *American Economic Review, 109*(8), 2889–2920. https://doi.org/10.1257/aer.20161574

Azoulay, P., Graff-Zivin, J., Uzzi, B., Wang, D., Williams, H., Evans, J. A., Jin, G. Z., Lu, S. F., Jones, B. F., Börner, K., Lakhani, K. R., Boudreau, K. J., & Guinan, E. C. (2018). Toward a more scientific science. *Science (New York, N.Y.), 361*(6408), 1194–1197. https://doi.org/10.1126/science.aav2484

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. *The RAND Journal of Economics, 42*(3), 527–554. https://doi.org/10.1111/j.1756-2171.2011.00140.x

Chen, Y., & Ding, J. (2023). Exploitation and exploration: An analysis of the research pattern of Nobel laureates in physics. *Journal of Informetrics, 17*(3), Article 101428. https://doi.org/10.1016/j.joi.2023.101428
Chu, J. S. G., & Evans, J. A. (2021). Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences, 118*(41), Article e2021636118. https://doi.org/10.1073/pnas.2021636118
Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., & Caldarelli, G. (2019). The statistical physics of real-world networks. *Nature Reviews Physics, 1*(1), 58–71. https://doi.org/10.1038/s42254-018-0002-6
Fortunato, S., Bergstrom, C. T., Boerner, K., Evans, J. A., Helbing, D., Milojevic, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabasi, A.-L. (2018). Science of science. *Science (New York, N.Y.), 359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185
Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' Research strategies. *American Sociological Review, 80*(5), 875–908. https://doi.org/10.1177/0003122415601618
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817. https://doi.org/10.1287/mnsc.2015.2366
Gates, A. J., Ke, Q., Varol, O., & Barabási, A. L. (2019). Nature's reach: Narrow work has broad impact. *Nature, 575*(7781), 32–34. https://doi.org/10.1038/d41586-019-03308-7
Guo, L., Wang, Y., & Li, M. (2024). Exploration, exploitation and funding success: Evidence from junior scientists supported by the Chinese Young Scientists Fund. *Journal of Informetrics, 18*(2), Article 101492. https://doi.org/10.1016/j.joi.2024.101492
Hill, R., & Stein, C. (2025). Race to the bottom: Competition and quality in science*. *The Quarterly Journal of Economics*, qjaf010. https://doi.org/10.1093/qje/qjaf010
Hill, R., Yin, Y., Stein, C., Wang, D., & Jones, B.F. (2021). Adaptability and the pivot penalty in science. https://doi.org/10.48550/arXiv.2107.06476.
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), Article 101409. https://doi.org/10.1016/j.joi.2023.101409
Huang, S., Lu, W., Bu, Y., & Huang, Y. (2022). Revisiting the exploration-exploitation behavior of scholars' research topic selection: Evidence from a large-scale bibliographic database. *Information Processing & Management, 59*(6), Article 103110. https://doi.org/10.1016/j.ipm.2022.103110
Huo, C., Ma, S., & Liu, X. (2022). Hotness prediction of scientific topics based on a bibliographic knowledge graph. *Information Processing & Management, 59*(4), Article 102980. https://doi.org/10.1016/j.ipm.2022.102980
Jia, T., Wang, D., & Szymanski, B. K. (2017). Quantifying patterns of research-interest evolution. *Nature Human Behaviour, 1*(4), 0078. https://doi.org/10.1038/s41562-017-0078
Lim, B. H., D'Ippoliti, C., Dominik, M., Hernández-Mondragón, A. C., Vermeir, K., Chong, K. K., Hussein, H., Morales-Salgado, V. S., Cloete, K. J., Kimengsi, J. N., Balboa, L., Mondello, S., dela Cruz, T. E., Lopez-Verges, S., Sidi Zakari, I., Simonyan, A., Palomo, I., Režek Jambrak, A., Germo Nzweundji, J., Molnar, A., Saktiawati, A. M. I., Elagroudy, S., Kumar, P., Enany, S., Narita, V., Backes, M., Siciliano, V., Egamberdieva, D., & Flores Bueso, Y. (2025). Regional and institutional trends in assessment for academic promotion. *Nature*. https://doi.org/10.1038/s41586-024-08422-9
Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature, 623*(7989), 987–991. https://doi.org/10.1038/s41586-023-06767-1
Liu, L., Dehmamy, N., Chown, J., Giles, C. L., & Wang, D. (2021). Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications, 12*(1), 5392. https://doi.org/10.1038/s41467-021-25477-8
Liu, L., Jones, B. F., Uzzi, B., & Wang, D. (2023). Data, measurement and empirical methods in the science of science. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01562-4
Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C., & Wang, D. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature, 559*(7714), 396–399. https://doi.org/10.1038/s41586-018-0315-8
Liu, M., Xie, Z., Yang, A. J., Yu, C., Xu, J., Ding, Y., & Bu, Y. (2024). The prominent and heterogeneous gender disparities in scientific novelty: Evidence from biomedical doctoral theses. *Information Processing & Management, 61*(4), Article 103743. https://doi.org/10.1016/j.ipm.2024.103743
Lu, C., Zhang, C., Xiao, C., & Ding, Y. (2022). Contributorship in scientific collaborations: The perspective of contribution-based byline orders. *Information Processing & Management, 59*(3), Article 102944. https://doi.org/10.1016/j.ipm.2022.102944
Ma, Y., & Uzzi, B. (2018). Scientific prize network predicts who pushes the boundaries of science. *Proceedings of the National Academy of Sciences, 115*(50), 12608–12615. https://doi.org/10.1073/pnas.1800485115
Manso, G., Balsmeier, B., & Fleming, L. (2023). Heterogeneous innovation over the business cycle. *The Review of Economics and Statistics, 105*(5), 1224–1236. https://doi.org/10.1162/rest_a_01104
Min, C., Bu, Y., Wu, D., Ding, Y., & Zhang, Y. (2021). Identifying citation patterns of scientific breakthroughs: A perspective of dynamic citation process. *Information Processing & Management, 58*(1), Article 102428. https://doi.org/10.1016/j.ipm.2020.102428
Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature, 613*(7942), 138–144. https://doi.org/10.1038/s41586-022-05543-x
Peng, H., Ke, Q., Budak, C., Romero, D. M., & Ahn, Y.-Y. (2021). Neural embeddings of scholarly periodicals reveal complex disciplinary organizations. *Science Advances, 7*(17), eabb9004. https://doi.org/10.1126/sciadv.abb9004
Savov, P., Jatowt, A., & Nielek, R. (2020). Identifying breakthrough scientific papers. *Information Processing & Management, 57*(2), Article 102168. https://doi.org/10.1016/j.ipm.2019.102168
Shi, D., Liu, W., & Wang, Y. (2023). Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists? *Science (New York, N.Y.), 379*(6627), 62–65. https://doi.org/10.1126/science.abq1218
Shi, F., & Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications, 14*(1), 1641. https://doi.org/10.1038/s41467-023-36741-4
Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks, 43*, 73–85. https://doi.org/10.1016/j.socnet.2015.02.006
Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science (New York, N.Y.), 354*(6312), aaf5239. https://doi.org/10.1126/science.aaf5239
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface, 4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science (New York, N.Y.), 342*(6157), 468–472. https://doi.org/10.1126/science.1240474
Van Buskirk, I., Clauset, A., & Larremore, D. B. (2023). An open-source cultural consensus approach to name-based gender classification. *Proceedings of the International AAAI Conference on Web and Social Media, 17*, 866–877. https://doi.org/10.1609/icwsm.v17i1.22195
Venturini, S., Sikdar, S., Rinaldi, F., Tudisco, F., & Fortunato, S. (2024). Collaboration and topic switches in science. *Scientific Reports, 14*(1), 1258. https://doi.org/10.1038/s41598-024-51606-6
Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies, 1*(1), 396–413. https://doi.org/10.1162/qss_a_00021
Wei, C., Li, J., & Shi, D. (2023). Quantifying revolutionary discoveries: Evidence from Nobel prize-winning papers. *Information Processing & Management, 60*(3), Article 103252. https://doi.org/10.1016/j.ipm.2022.103252
Wu, L. F., Wang, D. S., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378. https://doi.org/10.1038/s41586-019-0941-9. -+.
Xu, H., Liu, M., Bu, Y., Sun, S., Zhang, Y., Zhang, C., Acuna, D. E., Gray, S., Meyer, E., & Ding, Y. (2024). The impact of heterogeneous shared leadership in scientific teams. *Information Processing & Management, 61*(1), Article 103542. https://doi.org/10.1016/j.ipm.2023.103542

Yang, A. J. (2024). Unveiling the impact and dual innovation of funded research. *Journal of Informetrics, 18*(1), Article 101480. https://doi.org/10.1016/j.joi.2023.101480

Yang, A. J., Deng, S., Wang, H., Zhang, Y., & Yang, W. (2023a). Disruptive coefficient and 2-step disruptive coefficient: Novel measures for identifying vital nodes in complex networks. *Journal of Informetrics, 17*(3), Article 101411. https://doi.org/10.1016/j.joi.2023.101411

Yang, A. J., Gong, H., Wang, Y., Zhang, C., & Deng, S. (2024a). Rescaling the disruption index reveals the universality of disruption distributions in science. *Scientometrics, 129*(1), 561–580. https://doi.org/10.1007/s11192-023-04889-x

Yang, A. J., Hu, H., Zhao, Y., Wang, H., & Deng, S. (2023b). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management, 60*(5), Article 103420. https://doi.org/10.1016/j.ipm.2023.103420

Yang, A. J., Wu, L., Zhang, Q., Wang, H., & Deng, S. (2023c). The k-step h-index in citation networks at the paper, author, and institution levels. *Journal of Informetrics, 17*(4), Article 101456. https://doi.org/10.1016/j.joi.2023.101456

Yang, A. J., Xu, H., Ding, Y., & Liu, M. (2024b). Unveiling the dynamics of team age structure and its impact on scientific innovation. *Scientometrics*. https://doi.org/10.1007/s11192-024-04987-4

Yang, A. J., Yan, X., Hu, H., Hu, H., Kong, J., & Deng, S. (2025). Are disruptive papers more likely to impact technology and society? *Journal of the Association for Information Science and Technology, 76*(3), 563–579. https://doi.org/10.1002/asi.24947

Yang, A. J., Zhao, Y., Wang, H., & Deng, S. (2023d). Untangle the characteristics of disruptive and consolidating citations of Nobel-winning papers. In *86th Annual Meeting of the Association for Information Science and Technology*.

Yang, W., & Wang, Y. (2024). Higher-order structures of local collaboration networks are associated with individual scientific productivity. *EPJ Data Science, 13*(1), 15. https://doi.org/10.1140/epjds/s13688-024-00453-6

Yin, Y., Dong, Y., Wang, K., Wang, D., & Jones, B. F. (2022). Public use and public funding of science. *Nature Human Behaviour, 6*(10), 1344–1350. https://doi.org/10.1038/s41562-022-01397-5

Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E., & Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature Communications, 10*(1), 3439. https://doi.org/10.1038/s41467-019-11401-8

Zhai, S., Xia, L., Chi, M., & Li, X. (2024). Configurational determinants of time-to-win in NSFC youth program funding: Insights from Chinese library and information science. *Information Processing & Management, 61*(3), Article 103649. https://doi.org/10.1016/j.ipm.2024.103649

Zhang, X., Yin, D., Tang, L., & Zhao, H. (2024). Does academic engagement with industry come at a cost for early career scientists? Evidence from high-tech enterprises' Ph.D. Funding programs. *Information Processing & Management, 61*(3), Article 103669. https://doi.org/10.1016/j.ipm.2024.103669

Zhou, S., Chai, S., & Freeman, R. B. (2024). Gender homophily: In-group citation preferences and the gender disadvantage. *Research Policy, 53*(1), Article 104895. https://doi.org/10.1016/j.respol.2023.104895