The paradox of innovation: new concepts and disruptiveness in science

Abstract

This study examines the relationship between the introduction of new concepts and scientific disruptiveness using more than 42 million publications published between 1950 and 2024. We find that papers introducing new concepts are, on average, less disruptive than papers without new concepts. Historical analyses show that the relationship shifted from positive in earlier decades to consistently negative after the 1980s. We further show that team diversity strengthens the negative association between introducing new concepts and disruptiveness. Decomposition analyses indicate that papers introducing new concepts receive more citations overall, but disproportionately more consolidating citations than disruptive citations. Restricted cubic spline analyses reveal a statistically significant non-monotonic association. Low positive concentrations of new concepts are associated with lower disruptiveness, whereas the association reverses only in the sparsely populated upper tail of the distribution. Taken together, the findings reveal a complex relationship between conceptual change and scientific disruptiveness.

Keywords: Disruptiveness, Conceptual innovation, Science of science, Team diversity.

1 Introduction

Scientific progress depends on the continual introduction of new ideas, concepts, and research directions [1-7]. Throughout the history of science, major intellectual transformations have frequently been accompanied by the introduction of new terminology, from earlier concepts such as “DNA sequencing” and “nanotechnology” to more recent developments including “CRISPR,” “large language models (LLMs),” and “foundation models.” New concepts enable researchers to formalize emerging phenomena, communicate novel mechanisms, and structure previously disconnected bodies of knowledge. Consequently, the introduction of new terminology is often regarded as an important indicator of scientific innovation and intellectual change.

At the same time, a growing literature has emphasized that not all scientific innovation produces disruptive change [8-11]. Recent studies using large-scale citation data show that modern science has become increasingly consolidative, with fewer papers displacing prior knowledge and more papers integrating and extending existing research trajectories [9, 12]. This distinction between introducing something new and fundamentally reshaping science raises an important but understudied question: are papers that introduce new concepts more likely to become disruptive?

The answer is not obvious. On the one hand, introducing new concepts may facilitate disruptive discoveries by enabling researchers to challenge established paradigms and develop fundamentally new research directions [13, 14]. New terminology can signal the emergence of previously unrecognized phenomena, methods, or theoretical frameworks, thereby increasing the likelihood that subsequent work departs from existing knowledge structures. From this perspective, papers introducing new concepts should be more disruptive because they contribute to paradigm shifts and intellectual discontinuities. On the other hand, introducing new concepts may also reflect incremental specialization within increasingly complex scientific systems. Modern science produces a large volume of highly specialized terminology, much of which may correspond to narrow refinements rather than transformative advances. In such cases, new concepts may become integrated into existing knowledge frameworks instead of replacing them. Moreover, contemporary scientific production increasingly relies on large, diverse, and collaborative research teams, which may favor coordination, integration, and cumulative knowledge development over disruptive departures from prior work [8]. As a result, papers introducing new concepts may attract attention and citations while remaining relatively consolidative.

Despite extensive research on scientific novelty, disruptiveness, and innovation, little is known about the relationship between the introduction of new concepts and disruptive scientific change. Existing studies typically examine novelty through recombination of prior knowledge, atypical citations, or semantic distance measures, while separate streams of research analyze disruptiveness using citation network structures. Comparatively few studies directly investigate whether

the introduction of genuinely new scientific terminology is associated with greater disruptiveness. Furthermore, the historical evolution of this relationship remains poorly understood.

In this study, we address these questions using a large-scale analysis of more than 42 million scientific publications from the OpenAlex database spanning 1950–2024. Building on natural language processing frameworks, we identify newly introduced words and noun phrases from publication titles and abstracts to measure the introduction of new concepts. We then examine how these measures relate to scientific disruptiveness. Beyond the presence or absence of new concepts, the concentration of newly introduced terminology may also matter. Isolated new terms may be readily incorporated into established knowledge structures, whereas a paper organized around a high concentration of new concepts may reflect a broader departure from existing conceptual frameworks. We therefore also ask whether the association varies across the distribution of conceptual novelty and whether it can be adequately characterized by a linear functional form. We treat this question as exploratory rather than imposing a predetermined U-shaped or threshold model.

Our analysis generates several main findings. First, papers introducing new concepts are, on average, less disruptive than papers that do not introduce new concepts. This negative relationship is highly robust across scientific fields. Second, the relationship changes substantially over time. In earlier decades, introducing new concepts was positively associated with disruptiveness, whereas since the 1980s, the relationship has become consistently negative. Third, the association between new concepts and disruptiveness varies systematically with team diversity. Papers produced by more diverse teams—measured through institutional, international, age, and gender diversity—are less likely to introduce new concepts and are also less disruptive. Moreover, the negative relationship between introducing new concepts and disruptiveness becomes stronger within highly diverse teams. Fourth, decomposing the disruptiveness measure shows that papers introducing new concepts receive more citations overall, but disproportionately more consolidating citations than disruptive citations. Finally, formal spline analyses reveal a non-monotonic association between the concentration of new concepts and disruptiveness. Adjusted disruptiveness initially declines as the proportion of new concepts increases but subsequently rises and becomes positive relative to papers with no new concepts only in the upper tail of the distribution.

This study contributes to several strands of research. First, it distinguishes the textual introduction of new concepts from citation-based scientific disruption, showing that these two dimensions of innovation are not necessarily aligned. Second, it documents substantial historical and distributional heterogeneity in this association: the relationship changes over time and follows a non-monotonic functional form, with a positive fitted association appearing only in the sparsely populated upper tail. Third, it contributes to research on team science by showing that the association varies systematically across different dimensions of team diversity. These contributions are descriptive and associational; the analyses do not identify the causal processes producing the observed patterns.

2 Background

2.1 Disruptiveness

Scientific progress is fundamentally shaped by the tension between continuity and transformation, as articulated in foundational theories of knowledge evolution. Thomas Kuhn's paradigm shift framework [15] distinguishes between "normal science," which refines and consolidates existing paradigms through incremental advancements, and "revolutionary science," which disrupts established frameworks by introducing anomalies that lead to new paradigms. This duality underscores that impact extends beyond mere visibility—often proxied by citation counts [16, 17]—to the qualitative ways in which ideas either reinforce or redirect the trajectory of knowledge. Citation counts reveal frequency of reference but obscure whether a work bolsters prior foundations or overshadows them, potentially stifling attention to predecessors while opening novel avenues [2, 11].

Building on these theoretical insights, bibliometric approaches have sought to quantify disruptiveness as a measure of how scientific works alter knowledge flows. Disruptive innovations, akin to Schumpeter's creative destruction in economic theory, break from the past by rendering prior ideas less central, fostering paradigm shifts and new research domains [8-10]. In contrast, consolidating works strengthen existing paradigms, enabling cumulative progress by deepening ties to foundational knowledge. Despite limitations in fully capturing contextual influences [18], such metrics provide a lens for analyzing the structural role of ideas in reshaping fields, aligning with broader theories of innovation that emphasize the interplay between stability and upheaval [19-23]. Understanding disruptiveness is crucial not only for mapping scientific evolution but also for addressing observed declines in groundbreaking work, which may hinder societal advancements in technology and policy [8-10].

2.2 New Concepts

At the heart of scientific advancement lies the generation and dissemination of new ideas—empirical insights, theoretical models, or methodological tools—that either extend or challenge prevailing knowledge [24]. Drawing from recombination theories of innovation [25], new ideas often emerge from blending familiar elements with unconventional ones, balancing novelty with accessibility to overcome resistance from entrenched paradigms [26]. As Merton’s sociology

of science highlights [27], the adoption of ideas depends on their integration into the scientific community, where mechanisms like the Matthew effect amplify the influence of already prominent works, leading to uneven diffusion.

The impact of new ideas manifests in two interconnected phases: their initial introduction, which injects fresh perspectives into discourse, and their subsequent reuse, which signals broader acceptance and integration [28]. Interdisciplinary efforts exemplify this, as they merge disparate domains to yield transformative outcomes [29-31]. Beyond traditional citation-based assessments, substantive measures—such as tracking the emergence and propagation of novel concepts through textual analysis—offer a granular view of ideational influence [32]. These approaches reveal how ideas drive progress by either disrupting silos or consolidating fragmented knowledge, yet their relationship to disruptiveness remains theoretically underexplored. Motivated by the need to reconcile structural and content-based dimensions of impact, this integration can illuminate why certain ideas propel fields forward while others fade, informing strategies to counteract stagnation in innovation.

3 Methods and data

3.1 Data

This study uses the OpenAlex dataset [33], a large-scale open-source database of global scholarly publications maintained by OurResearch at OpenAlex. OpenAlex integrates information from multiple sources, including Microsoft Academic Graph (MAG), Crossref, PubMed, ORCID, and publisher websites, to provide a standardized database of more than 250 million scholarly works published between 1666 and 2025. The dataset used in this study was downloaded in February 2025.

OpenAlex includes a wide range of scholarly outputs, such as journal articles, conference papers, books, datasets, and dissertations. It also records detailed relationships among works, authors, institutions, journals, citations, concepts, and funders. One important advantage of OpenAlex is its entity disambiguation system, which assigns unique identifiers to authors, works, and institutions using information such as author names, publication histories, citation patterns, and ORCID records.

The version of OpenAlex used in this study adopts a four-level classification system consisting of domains, fields, subfields, and topics. At the highest level, the dataset contains four domains: Health Sciences, Life Sciences, Social Sciences, and Physical Sciences. These domains are further divided into 26 fields, such as Medicine and Mathematics, 252 subfields, and 4,516 topics. Topics are assigned automatically based on publication titles, abstracts, journal names, and citation relationships. Each work is classified according to its highest-scoring topic, which determines its corresponding subfield, field, and domain. In the empirical analysis, fields are included as fixed effects, while the effects of new ideas are examined at both the field and domain levels.

To improve data quality, we follow the data-cleaning procedure proposed by Arts, Melluso [28]. First, we restrict the sample to non-retracted English-language journal articles and conference proceedings by selecting works with "primary_location" source types equal to "journal" or "conference," language equal to "en," and "is_retracted" equal to "false." We further exclude publications from journals without publishers or with an Impact Factor of zero. Publications with duplicate titles or abstracts are removed, retaining only one observation in each case. We also exclude publications with missing titles or without listed authors.

After applying these filters, the dataset contains 75,612,702 publications published between 1666 and 2023. For the citation and disruptiveness analyses, citation data are included through the end of 2024 to ensure at least a one-year citation window. We further restrict the sample to papers published between 1950 and 2024 that have at least one citation and one reference. The final analytical sample consists of 42,566,298 papers.

3.2 Quantifying disruptiveness

To measure scientific disruptiveness, we use the CD index, a widely adopted indicator originally introduced by Funk and Owen-Smith [11]. The CD index evaluates whether a paper disrupts existing knowledge or consolidates prior research by examining its local citation network, including its references, citing papers, and citation relationships among them.

The underlying idea is that a disruptive paper reduces the future relevance of its predecessors. If later studies cite the focal paper without also citing its references, the focal paper is considered disruptive because it replaces earlier work as the primary source of knowledge. In contrast, if later studies continue to cite both the focal paper and its predecessors, the focal paper is considered consolidating because it reinforces existing knowledge.

$$CD_\tau = \frac{1}{n^\tau}\sum_{i=1}^{n}\frac{-2f_i^\tau b_i^\tau + f_i^\tau}{w_i^\tau},$$

where $f_i^\tau = 1$ if paper i cites the focal paper (FP), else 0; $b_i^\tau = 1$ if paper i cites any of the FP's predecessors, else 0.

Here, n^τ represents the sum of papers that cite: (1) only the focal paper (type f), (2) both the focal paper and parts of its predecessors (type b), or (3) only parts of its predecessors. w_i^τ indexes a matrix of weights assigned to papers. A citation increases a paper's consolidating nature if the citing paper also cites the FP's predecessors, or its disruptive nature if it does not. The CD index ranges from −1 (consolidating) to +1 (disruptive). A lower CD index indicates alignment with established knowledge, while a higher value suggests potential to disrupt scientific paradigms. Although the CD index assumes a paper's references form its knowledge base and can be affected by reference count, its simplicity makes it widely used in citation-based analyses.

A forward citation increases the disruptive nature of a focal paper when the citing paper cites the focal paper but not its predecessors. Conversely, a citation increases the consolidating nature of the focal paper when the citing paper cites both the focal paper and its predecessors. The CD index ranges from −1 to +1. Values close to −1 indicate highly consolidating research that builds upon established knowledge, whereas values close to +1 indicate highly disruptive research with greater potential to shift scientific paradigms.

Although the CD index assumes that a paper's references adequately represent its knowledge base and may be influenced by the number of references, it remains one of the most widely used measures of scientific disruptiveness because of its conceptual clarity and computational simplicity.

Because the distribution of CD indices is highly non-normal, we follow Wu et al. [8] and transform the raw CD scores into percentile-based disruptiveness values. When multiple papers share the same CD index, we assign them the median percentile rank.

3.3 Quantifying new concepts

New concepts are fundamental to scientific progress because they enable researchers to develop new theories, methods, and research directions. To measure new concepts, we follow the natural language processing framework developed by Arts, Melluso [28], which identifies the emergence of new scientific terminology within large-scale publication corpora.

Our analysis is based on publication metadata retrieved from the OpenAlex database, including titles, abstracts, publication years, authorship information, affiliations, and citation links. Following prior work, we combine publication titles and abstracts into a unified text corpus for analysis. Because OpenAlex does not provide comprehensive full-text coverage, the analysis relies on titles and abstracts rather than full texts. Approximately 20% of publications in the database do not contain abstracts and therefore rely only on title information. We acknowledge that this may limit the ability to fully capture the conceptual content of some publications, particularly when important scientific contributions are not fully reflected in abstracts.

The text preprocessing pipeline follows standard NLP procedures implemented primarily in Python using libraries including spaCy and NLTK. Specifically, all text is converted to lowercase and tokenized into individual terms. We then apply part-of-speech tagging and dependency parsing to identify noun phrases and multi-word scientific expressions. Lemmatization is used to reduce words to their base forms, while stop-word removal and text cleaning procedures eliminate punctuation, non-informative tokens, and generic vocabulary. Dependency-based chunk extraction is further used to identify scientifically meaningful noun phrases and conceptual combinations.

Using this framework, a new word is defined as a unique unigram that appears for the first time in the historical publication record, while a new phrase is defined as a newly appearing noun phrase or multi-word scientific expression. To reduce noise arising from typographical errors or highly idiosyncratic expressions, we exclude terms that appear only once in the full corpus. We then construct binary indicators identifying whether a publication introduces at least one new word or one new phrase. The empirical analyses in this study focus on publications published between 1950 and 2024. Earlier historical records are used only to establish the baseline vocabulary necessary for identifying genuinely new concepts over time.

Building on this methodology, we construct measures of new concepts. First, new word is a dummy variable indicating whether a paper introduces at least one newly appearing unigram. Second, new phrase is a dummy variable indicating whether a paper introduces at least one newly appearing noun phrase, which we interpret as a new concept. These two dummy measures serve as our primary indicators of conceptual novelty throughout the analysis. In addition, we construct two continuous measures that capture the intensity of conceptual novelty within a paper. Specifically, proportion of new words measures the share of newly introduced unigrams among all extracted unigrams in a publication, while proportion of new phrases measures the share of newly introduced noun phrases among all extracted noun phrases.

We primarily rely on the new phrase measure because many important scientific advances are expressed through multi-word expressions rather than isolated words. Examples such as "deep learning," "gene editing," and "graph neural

networks" illustrate how new combinations of words often represent new ideas, methods, or phenomena. By focusing on newly emerging phrases, this measure more effectively captures the introduction of novel conceptual structures in scientific research.

Although this approach is not without limitations, it offers several important advantages. First, it enables the large-scale and systematic measurement of conceptual novelty across millions of publications over long historical periods. Second, scientific phrases closely reflect how researchers communicate and formalize new ideas in the literature. Third, the first appearance of a new scientific word or phrase often signals the emergence of a distinct research direction or conceptual innovation.

3.4 Empirical strategy

To examine the relationship between the introduction of new concepts and scientific disruptiveness, we estimate a series of ordinary least squares (OLS) regressions in which paper disruptiveness is regressed on indicators capturing the introduction of new words and new phrases. The baseline specification is given by:

$$Disruptiveness_i = \beta_0 + \beta_1 NewConcepts_i + \gamma X_i + \delta_t + \phi_f + \varepsilon_i,$$

where $Disruptiveness_i$ represents the disruptiveness percentile of paper i, $NewConcepts_i$ denotes our measures of the introduction of new concepts, X_i is a vector of control variables, δ_t represents publication year fixed effects, ϕ_f denotes field fixed effects, and ε_i is the error term.

Our primary explanatory variables are the dummy indicators for new word and new phrase, which capture whether a publication introduces at least one newly appearing unigram or noun phrase. In additional analyses, we also use the proportion of new words and proportion of new phrases, both ranging from 0 to 1, to measure the extent to which a publication relies on newly introduced terminology.

The regressions include several controls related to team composition, author characteristics, and publication attributes. Specifically, we control for team size, number of references, first author citation count, first author career age, and first author productivity. To avoid simultaneity concerns, author citation counts and productivity are calculated using only publications and citations accumulated prior to the focal paper. First author productivity is measured by the total number of publications produced before the focal publication year, while first author citation count is measured by cumulative citations received before publication of the focal paper.

We also include a dummy variable indicating whether the first author is female. Author gender is estimated using first names based on name-to-gender inference algorithms [34]. In addition, we control for first author institutional tier, categorized into four groups according to institutional citation rankings: top 1%, top 1–5%, top 5–20%, and all remaining institutions. Author affiliation, institutional information, and country information are obtained from the OpenAlex dataset. Institutional affiliations are based on disambiguated organization identifiers provided by OpenAlex, while country information is derived from authors' affiliated institutions at the time of publication.

Publication year fixed effects account for temporal variation in publication practices, citation behavior, and scientific development, while field fixed effects control for disciplinary differences in citation patterns and research organization.

To assess whether the association varies with team composition, we examine several dimensions of research-team diversity. First, we measure multi-institutional collaboration using the number of unique disambiguated institutions represented among coauthors [35]. Second, we measure international collaboration using the number of distinct countries represented within a research team [36]. We also examine diversity within research teams. Team age diversity is measured using entropy based on authors' career ages [37], while team gender diversity is measured using entropy based on the gender composition of coauthors [38]. These analyses allow us to explore whether collaboration patterns and team diversity help explain the relationship between the introduction of new concepts and scientific disruptiveness.

To assess the functional relationship between the concentration of new concepts and disruptiveness without imposing arbitrarily selected cutoffs, we estimate restricted cubic spline models for the proportions of new words and new phrases. Because both measures contain a substantial mass at zero, each model includes a separate indicator distinguishing papers with zero new concepts from papers with positive proportions. The spline function is estimated over the positive-value distribution.

We use four knots located at the 5th, 35th, 65th, and 95th percentiles of the positive-value distribution. The corresponding knot values are 0.0127, 0.0208, 0.0317, and 0.1111 for the proportion of new words, and 0.0286, 0.0526, 0.0909, and 0.3333 for the proportion of new phrases. All models include the same control variables and publication-year and field fixed effects as the baseline models.

We formally evaluate nonlinearity using joint Wald tests of the nonlinear spline terms. Adjusted predictions are expressed relative to papers with zero new-concept proportions. Because the locations of curve minima and zero crossings

are obtained from the fitted spline rather than from a structural-break model, we interpret them as approximate turning regions rather than formally identified thresholds.

4 Results

4.1 New concepts are associated with lower disruptiveness

Our main objective is to examine the relationship between the introduction of new concepts and scientific disruptiveness. The introduction of new words and phrases represents an ex ante form of scientific change because it captures the appearance of new terminology or conceptual combinations at the time of publication. In contrast, disruptiveness is an ex post outcome that reflects how subsequent research responds to a paper through citation behavior. To study this relationship, we construct binary indicators for both new word and new phrase. A paper is assigned a value of 1 if it introduces at least one newly appearing word or phrase, and 0 otherwise.

Figure 1a–b compares the distributions of disruptiveness between papers that do and do not introduce new concepts. We find that papers without new words have a mean disruptiveness percentile of 0.501, whereas papers introducing new words have a lower mean value of 0.484. A similar pattern appears for new phrases. Papers without new phrases have a mean disruptiveness percentile of 0.510, compared with 0.475 for papers introducing new phrases. Although the differences are moderate in magnitude, the results reveal a notable contrast between the introduction of new concepts and subsequent scientific influence: papers that introduce new terminology or conceptual combinations tend to be less disruptive in later scientific development.

This pattern persists across different team sizes. Figure 1c–d repeats the analysis separately for research teams of different sizes. Consistent with prior studies, larger teams generally produce less disruptive research. However, within every team-size category, papers introducing new words or new phrases continue to exhibit lower disruptiveness than papers without them. This suggests that the negative relationship between the introduction of new concepts and disruptiveness is not solely driven by differences in team composition.

To further examine this relationship, we estimate ordinary least squares (OLS) regressions predicting disruptiveness using indicators for new words and new phrases. The regressions include controls for team size, number of references, first author citation count, first author career age, first author productivity, first author gender, and institutional tier. We also include publication year and field fixed effects to account for temporal and disciplinary variation in publication practices and citation behavior.

The regression results, presented in Figure 1e–f, show consistently negative and statistically significant coefficients for both new words (coef. = −1.207, $p<0.001$) and new phrases (coef. = −2.194, $p<0.001$). These findings indicate that papers introducing new terminology or new conceptual combinations are, on average, less disruptive in subsequent scientific development.

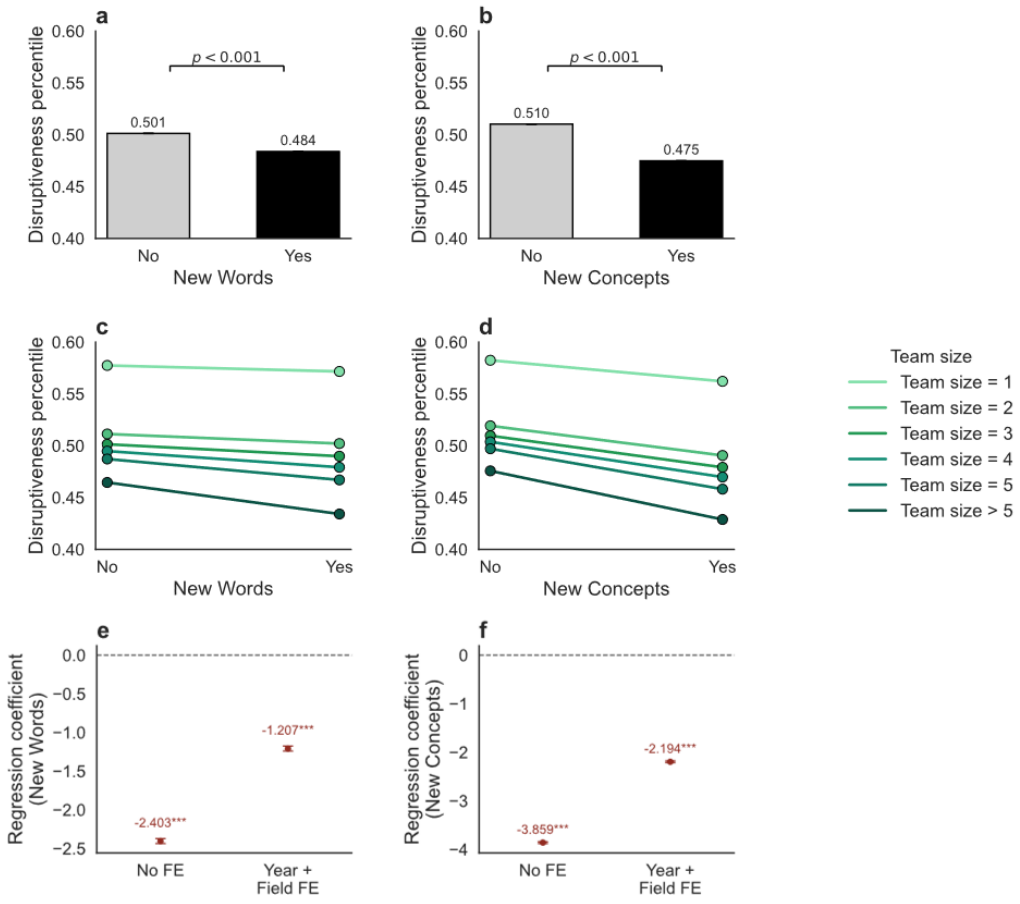


Fig. 1. Ex ante–ex post innovation paradox: new concepts are associated with lower disruptiveness.

We further examine whether this relationship varies across scientific disciplines by estimating separate OLS regressions for each of the 26 fields in the OpenAlex classification system. The results reveal a highly consistent pattern across disciplines. For the new word measure, the estimated coefficients are negative in 23 of the 26 fields. For the new phrase measure, all 26 fields exhibit negative and statistically significant coefficients.

Overall, these findings demonstrate that the negative association between the introduction of new concepts and scientific disruptiveness is highly robust across scientific domains and is not driven by a small subset of disciplines.

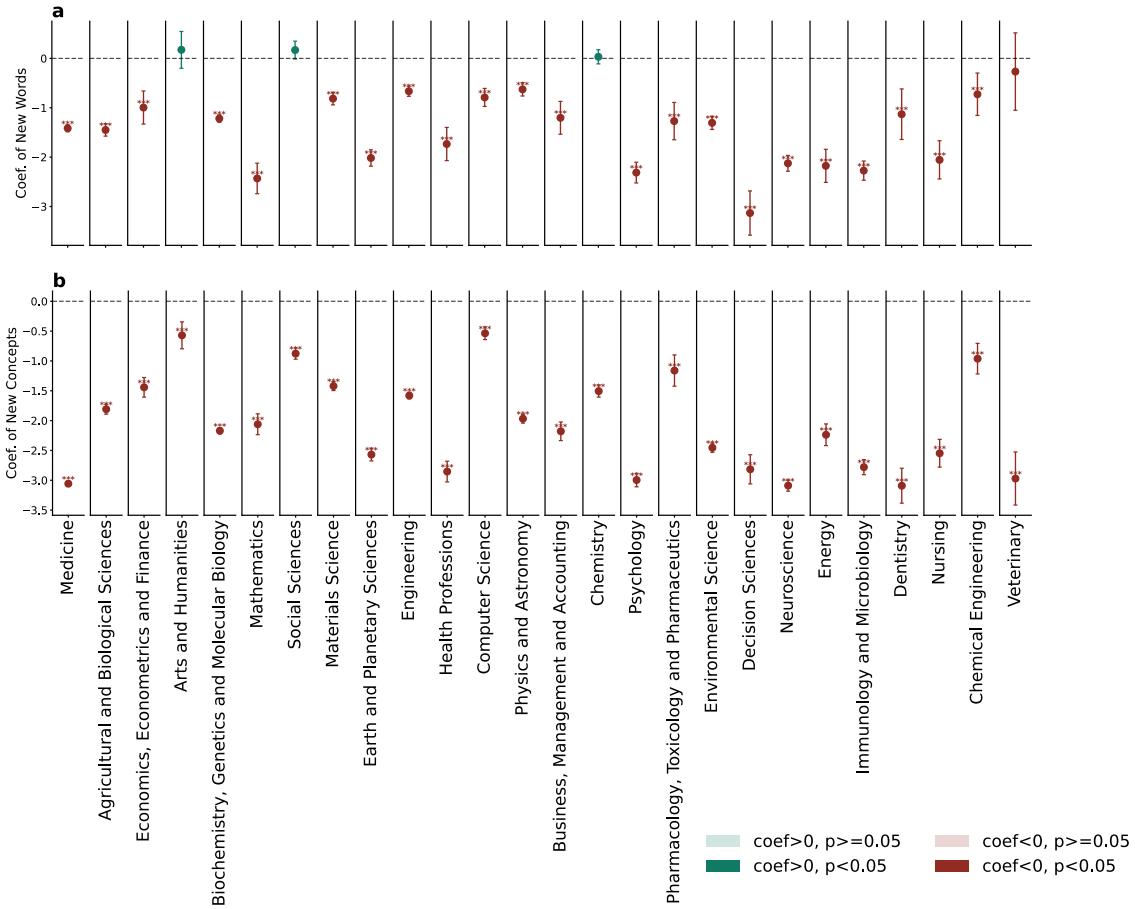


Fig. 2 Negative association between new concepts and disruptiveness across scientific fields.

4.2 Temporal dynamics of the relationship between new concepts and disruptiveness

To examine whether the relationship between the introduction of new concepts and disruptiveness changes over time, we estimate separate regression models for papers published in each year. This approach allows us to trace how the association between new concepts and disruptiveness evolves across historical periods.

Fig. 3 shows that the negative relationship observed in the full sample is not historically stable. In earlier decades, the relationship between the introduction of new concepts and disruptiveness was often positive. For example, in 1967, the coefficient for new words in predicting disruptiveness percentile is 1.494 and statistically significant at the 0.001 level. Similarly, the coefficient for new phrases in 1967 is 1.131 and also statistically significant at the 0.001 level. These results suggest that papers introducing new concepts during this period were more likely to become disruptive.

However, this pattern gradually reverses over time. Beginning around the 1980s and becoming particularly pronounced after 1990, the yearly coefficients for both new words and new phrases become consistently negative and statistically significant. In recent decades, papers introducing new concepts are systematically associated with lower disruptiveness.

These findings suggest that the role of new concepts in scientific development has changed substantially over time. In earlier periods, the introduction of new concepts may have reflected more fundamental conceptual breakthroughs. Scientific terminology was less saturated, and the emergence of a genuinely new term or phrase may have corresponded more directly to the formation of a new research direction or scientific paradigm. In contrast, contemporary scientific publications may generate a larger number of incremental, highly specialized, or narrowly defined terms that do not necessarily represent major conceptual advances. As a result, newly introduced terminology in recent decades may capture linguistic differentiation more often than transformative scientific change.

The observed trend may also reflect broader structural changes in the organization of science. Prior research has documented a long-term decline in scientific disruptiveness, with modern science becoming increasingly incremental and consolidative [10]. At the same time, the rapid expansion of the scientific literature has altered the dynamics of introducing new concepts [39]. In a more mature and densely interconnected scientific system, papers introducing new concepts may be more likely to extend or integrate existing knowledge rather than replace prior research, thereby generating lower disruptiveness.

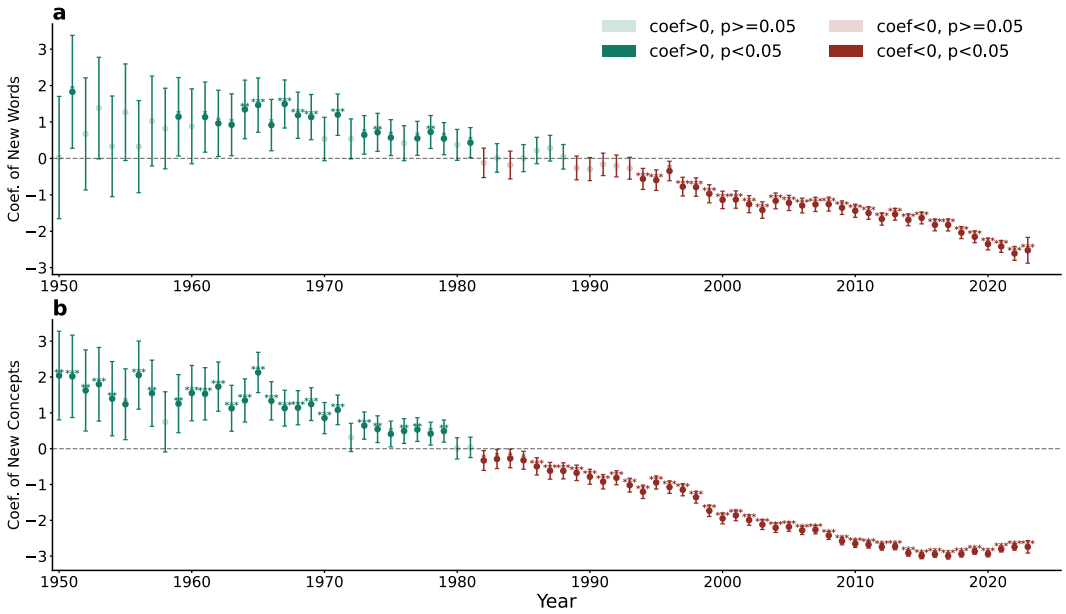


Fig. 3 Changing association between new concepts and disruptiveness over time.

4.3 Heterogeneity in the association by team diversity

To examine heterogeneity in the association, we analyze four dimensions of team diversity. Specifically, we analyze four dimensions of team diversity: multi-institutional collaboration, international collaboration, age diversity, and gender diversity.

Multi-institutional collaboration is measured by the number of unique institutions represented among coauthors, while international collaboration is measured by the number of distinct countries represented within a team. Age diversity is calculated using entropy based on authors' career ages, and gender diversity is calculated using entropy based on the gender composition of team members.

We first divide the sample into high-diversity and low-diversity groups based on the median value of each diversity measure. We then compare disruptiveness and the probability of introducing new concepts across these groups. Figure 4 shows a consistent pattern across all four dimensions of diversity: more diverse teams tend to produce less disruptive research and are also less likely to introduce new concepts.

For multi-institutional collaboration, papers produced by highly multi-institutional teams exhibit significantly lower disruptiveness percentiles than papers produced by less multi-institutional teams (0.471 vs. 0.527, $p<0.001$). These teams are also less likely to introduce new words (0.068 vs. 0.079, $p<0.001$) and new phrases (0.279 vs. 0.302, $p<0.001$). A similar pattern emerges for international collaboration. Highly international teams produce papers with lower disruptiveness percentiles (0.452 vs. 0.512, $p<0.001$), lower probabilities of introducing new words (0.071 vs. 0.074, $p<0.001$), and lower probabilities of introducing new phrases (0.285 vs. 0.292, $p<0.001$). Teams with greater age diversity also exhibit lower disruptiveness (0.469 vs. 0.526, $p<0.001$), as well as lower probabilities of introducing new words (0.073 vs. 0.074, $p<0.001$) and new phrases (0.289 vs. 0.292, $p<0.001$). Likewise, teams with higher gender diversity produce less disruptive research (0.481 vs. 0.516, $p<0.001$) and are less likely to introduce new words (0.070 vs. 0.077, $p<0.001$) and new phrases (0.278 vs. 0.301, $p<0.001$).

These results suggest that highly diverse teams are less likely to generate disruptive scientific change or introduce new conceptual combinations [9, 19, 40]. One possible explanation is that diverse teams face higher coordination costs and stronger incentives to build consensus across heterogeneous members, institutions, and disciplinary backgrounds. Such collaboration structures may encourage more incremental and consolidative forms of research rather than highly disruptive work.

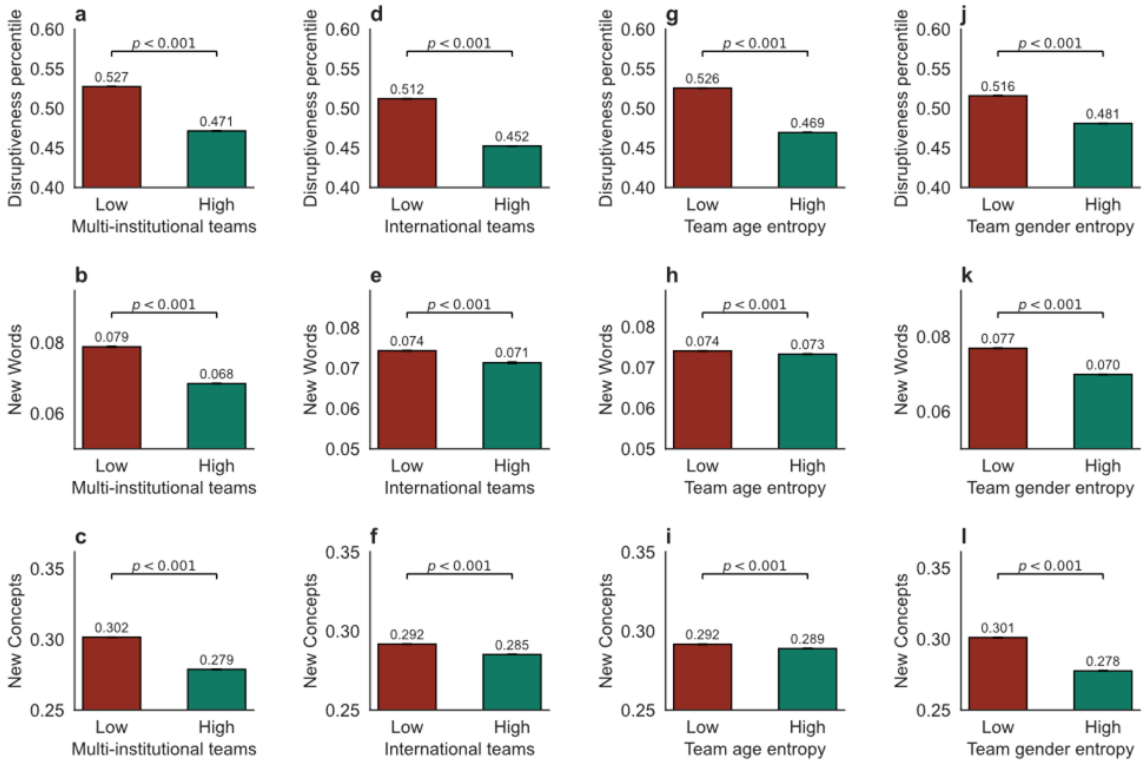


Fig. 4 Team diversity, introduction of new concepts, and disruptiveness.

To further assess this heterogeneity, we incorporate interaction terms between the team-diversity measures and the introduction of new concepts. Table 1 shows that all interaction terms are negative and statistically significant, indicating that the negative relationship between introducing new concepts and disruptiveness becomes stronger in more diverse teams.

For new words, the interaction coefficients are negative for multi-institutional collaboration ($\beta=-1.611$, $p<0.001$), international collaboration ($\beta=-2.232$, $p<0.001$), age diversity ($\beta=-0.590$, $p<0.001$), and gender diversity ($\beta=-1.479$, $p<0.001$). Similarly, for new concepts, the interaction coefficients are also consistently negative: multi-institutional collaboration ($\beta=-1.768$, $p<0.001$), international collaboration ($\beta=-2.098$, $p<0.001$), age diversity ($\beta=-1.010$, $p<0.001$), and gender diversity ($\beta=-1.619$, $p<0.001$).

The interaction results indicate that team diversity moderates the association between the introduction of new concepts and disruptiveness. Although diverse teams may possess broader knowledge resources and perspectives, they may also rely more heavily on coordination, integration, and consensus-building processes that favor incremental extensions of existing knowledge over disruptive departures from prior work. As a result, newly introduced concepts emerging from highly diverse teams may be more likely to become incorporated into established scientific frameworks rather than displacing them. These organizational interpretations are consistent with the observed interaction patterns but are not directly tested by the present design.

Table 1. Interaction effects of team diversity and introduction of new concepts on disruptiveness.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent Var.				**Disruptiveness percentile**				
Independent Var.								
New Words	0.4054*** (0.0422)	0.5180*** (0.0478)	-0.5390*** (0.0317)	-0.8710*** (0.0229)				
Multi-institution	-1.648*** (0.0129)				-1.276*** (0.0140)			
New Words ×Multi-institution	-1.611*** (0.0389)							
International		-3.747*** (0.0179)				-3.291*** (0.0200)		
New Words ×International		-2.232*** (0.0589)						
Age Diversity			-5.044*** (0.0189)				-4.744*** (0.0192)	

New Words × Age Diversity			-0.5900*** (0.0242)					
Gender Diversity				0.5931*** (0.0173)				0.9579*** (0.0190)
New Words × Gender Diversity				-1.479*** (0.0529)				
New Concepts					-0.4023*** (0.0246)	-0.5569*** (0.0284)	-1.027*** (0.0186)	-1.888*** (0.0134)
New Concepts ×Multi-institutional					-1.768*** (0.0225)			
New Concepts ×International						-2.098*** (0.0347)		
New Concepts × Age Diversity							-1.010*** (0.0142)	
New Concepts × Gender Diversity								-1.619*** (0.0303)
Control Var.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42,566,298	42,566,298	42,566,298	38,815,354	42,566,298	42,566,298	42,566,298	38,815,354
Adj. R^2	0.06348	0.06417	0.06464	0.06395	0.06462	0.06524	0.06575	0.06526

Note: Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.4 Decomposing disruptiveness

To better understand why the introduction of new concepts is associated with lower disruptiveness, we further decompose the disruptiveness measure into its underlying citation components. Specifically, we separate forward citations into disruptive citations and consolidating citations based on whether the citing paper also cites the references of the focal paper. A forward citation is classified as disruptive if the citing paper cites the focal paper without citing its predecessors, whereas it is classified as consolidating if the citing paper cites both the focal paper and its references. In addition, we measure the total citations received by the references of the focal paper as an indicator of knowledge-base impact.

The results, reported in Table 2, show that introducing new words and new phrases significantly increases both disruptive citations and consolidating citations. These findings suggest that papers introducing new concepts generally receive greater scientific attention and have higher overall scientific impact.

However, the positive effect is substantially larger for consolidating citations than for disruptive citations. For new words, the estimated coefficient is 0.0924 for disruptive citations compared with 0.1522 for consolidating citations. Similarly, for new phrases, the coefficient is 0.1672 for disruptive citations and 0.2411 for consolidating citations. These results indicate that papers introducing new concepts are more likely to stimulate follow-up research that continues to engage with the existing knowledge base rather than replacing it.

This imbalance helps explain why the introduction of new concepts is associated with lower overall disruptiveness. Although papers introducing new concepts attract more citations overall, a larger share of these citations are consolidating rather than disruptive. As a result, the relative proportion of disruptive citations declines, leading to lower disruptiveness scores. In other words, newly introduced concepts tend to become integrated into existing scientific frameworks instead of displacing prior knowledge.

Table 2. Decomposition of the relationship between new concepts and disruptiveness.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent Var.	**ln (Disruptive citations)**		**ln (Consolidating citations)**		**ln (Knowledge base impact)**	
Independent Var.						
New Words	0.0924*** (0.0006)		0.1522*** (0.0006)		0.0661*** (0.0008)	
New Concepts		0.1672*** (0.0004)		0.2411*** (0.0004)		0.0461*** (0.0005)
Control Var.	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42,566,298	42,566,298	42,566,298	42,566,298	42,566,298	42,566,298
Adj. R^2	0.05521	0.05902	0.29030	0.29674	0.41683	0.41686

Note: Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.5 Nonlinearity and upper-tail reversals in the association

The preceding analyses focus primarily on whether a paper introduces at least one new word or new phrase. However, the association between conceptual novelty and disruptiveness may also depend on the concentration of new concepts within a paper. We therefore examine the proportions of new words and new phrases, defined as the numbers of newly introduced words and phrases divided by the corresponding total numbers extracted from each publication.

To assess the functional form without imposing arbitrarily selected cutoffs, we estimate restricted cubic spline models. Because both proportion measures contain a substantial mass at zero, the models distinguish papers with no new concepts from papers with positive proportions and estimate the spline functions over the positive-value distributions. All models include the same control variables and publication-year and field fixed effects as the baseline regressions.

Figure 5a–b presents the adjusted spline estimates relative to papers with no new words or new phrases. Joint Wald tests strongly reject the null hypothesis that the nonlinear spline components are jointly equal to zero. The test statistic is 189.9 for the proportion of new words and 1,024.7 for the proportion of new phrases, with $p<0.001$ in both cases. These results provide formal evidence that the associations cannot be adequately characterized by a simple linear specification.

For new words, the adjusted association initially declines as the proportion of new words increases. The fitted curve reaches its approximate minimum at a proportion of 0.024, where predicted disruptiveness is about 2.4 percentile points lower than that of papers with no new words. The association subsequently increases and crosses the zero-reference line at approximately 0.112. Beyond this point, papers with higher proportions of new words are predicted to be more disruptive than papers containing no new words.

A similar but more gradual pattern is observed for new phrases. Predicted disruptiveness initially decreases and reaches its approximate minimum at a new-phrase proportion of 0.062, corresponding to a reduction of approximately 3.7 percentile points relative to papers with no new phrases. The fitted association then rises and crosses the zero-reference line at approximately 0.503. Thus, the positive association emerges only when newly introduced phrases constitute an exceptionally large share of the conceptual content of a paper.

The descriptive relationships shown in Figure 5c–d are consistent with the spline estimates. Papers with low positive proportions of new concepts generally exhibit lower disruptiveness, whereas disruptiveness increases among papers located further into the upper tail. However, the confidence intervals in the most extreme bins become wider, reflecting the limited number of observations at very high proportions of new concepts.

The distributions in Figure 5e–f further demonstrate that the upper-tail reversal applies to only a small minority of publications. Approximately 0.329% of papers lie beyond the estimated zero-crossing point for new words, while approximately 0.263% lie beyond the corresponding point for new phrases. Most publications either contain no new concepts or contain them only in relatively low proportions. Consequently, the positive association observed at high concentrations of new concepts is substantively notable but empirically rare.

Taken together, these results indicate a statistically significant non-monotonic association between the concentration of new concepts and scientific disruptiveness. Introducing a limited proportion of new terminology is associated with lower disruptiveness, whereas the association reverses among papers located in the sparsely populated upper tail of the distribution. This pattern suggests that isolated additions to the scientific vocabulary may generally be incorporated into existing knowledge structures, while unusually concentrated conceptual departure is more strongly associated with disruptive research.

Nevertheless, the approximate minima and zero-crossing points are derived from fitted spline curves and should not be interpreted as universal or structurally identified thresholds. Their precise locations may depend on the model specification, sample composition, and the density of observations in the upper tail. We therefore characterize the finding as an upper-tail reversal in a non-monotonic association, rather than as evidence of a deterministic threshold at which new concepts necessarily become disruptive.

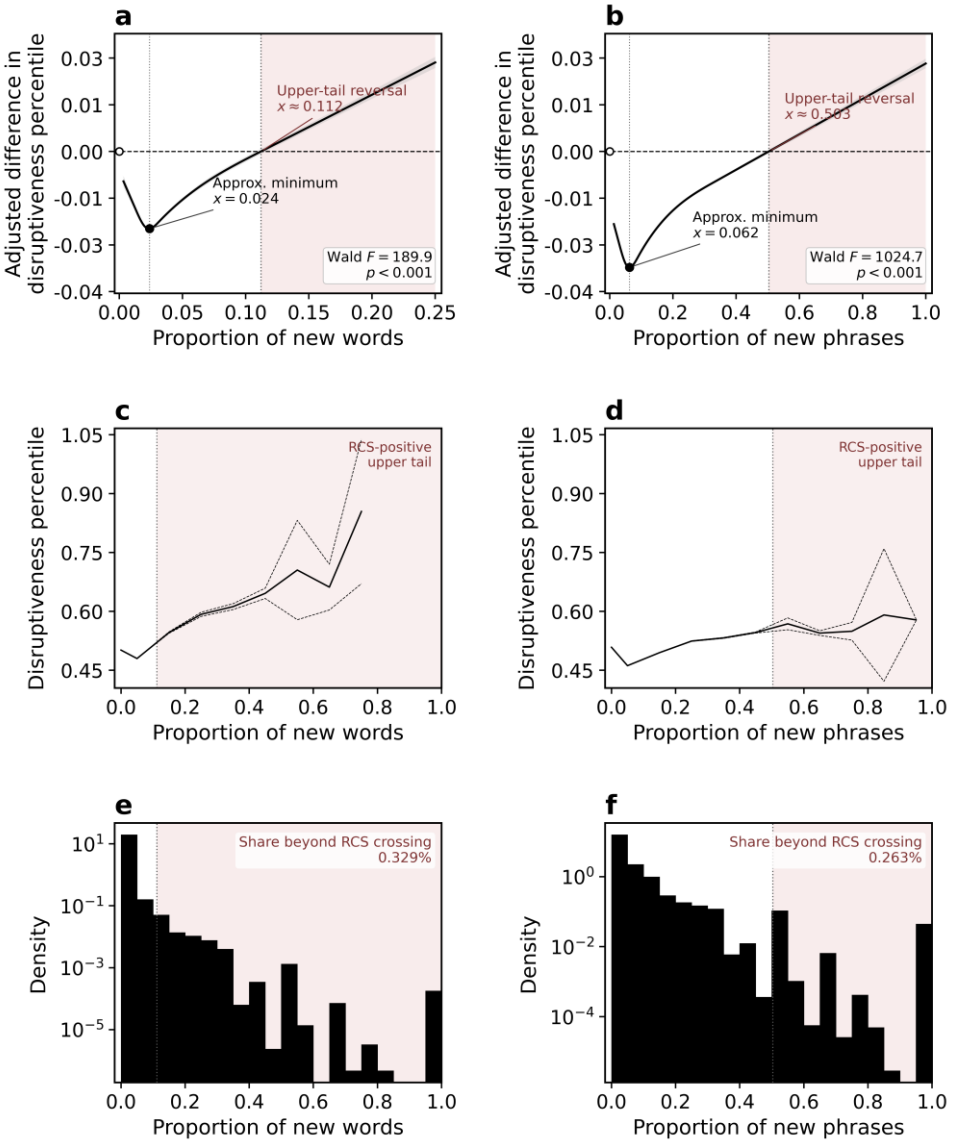


Fig. 5 Restricted cubic spline estimates of the association between the proportion of new concepts and scientific disruptiveness. Panels a and b present adjusted estimates from restricted cubic spline models for the proportions of new words and new phrases, respectively. The solid lines show the fitted associations relative to papers with zero new-concept proportions, and the gray bands indicate 95% confidence intervals. The horizontal dashed line denotes zero adjusted difference. The vertical dotted lines identify the approximate fitted minimum and the point at which the fitted association crosses zero. The shaded areas indicate the upper-tail regions in which the fitted associations are positive relative to papers with no new concepts. Joint Wald tests evaluate whether the nonlinear spline terms are jointly equal to zero. Panels c and d show the corresponding descriptive relationships based on binned new-concept proportions; solid lines indicate bin means and dashed lines indicate 95% confidence intervals. Panels e and f show the distributions of the proportions of new words and new phrases, with density displayed on a logarithmic scale. The percentages reported in these panels indicate the shares of observations beyond the corresponding RCS zero-crossing points. All regression models include the full set of control variables and publication-year and field fixed effects. The fitted minima and zero-crossing points are model-based summaries and should not be interpreted as structurally identified thresholds.

Table 3 summarizes the formal tests and key features of the restricted cubic spline estimates. The analyses are based on 42,566,298 observations for new words and 42,173,026 observations for new phrases. Joint Wald tests strongly reject the null hypothesis that the nonlinear spline terms are jointly equal to zero. The Wald statistic is 189.93 for new words and 1,024.66 for new phrases, with two nonlinear degrees of freedom and $p<0.001$ in both models. These results confirm that a linear specification is insufficient to characterize the association between the concentration of new concepts and disruptiveness.

The fitted curves also reveal substantively meaningful turning patterns. For new words, the estimated association reaches its approximate minimum at a proportion of 0.0239, where predicted disruptiveness is 2.479 percentile points lower than that of papers with no new words (95% CI: −2.526 to −2.431). The fitted association subsequently rises and crosses the zero-reference line at approximately 0.112. The lower bound of the 95% confidence interval becomes positive at approximately 0.116, indicating that the association is statistically positive beyond this point within the fitted curve.

For new phrases, the approximate minimum occurs at a proportion of 0.0624, corresponding to a reduction of 3.723 percentile points relative to papers with no new phrases (95% CI: −3.749 to −3.696). The fitted curve crosses the zero-reference line at approximately 0.503, while the lower bound of the 95% confidence interval becomes positive at approximately 0.514. Thus, the reversal occurs considerably further into the distribution for new phrases than for new words.

Importantly, the approximate minima, zero-crossing points, and confidence-bound crossing points are derived from the fitted spline curves rather than from a formal structural-break estimator. They should therefore be interpreted as descriptive turning regions within the estimated functional form, rather than as universal or precisely identified thresholds.

Table 3. Formal tests and summary measures of nonlinear associations between new-concept concentration and disruptiveness.

	New words	New phrases
Observations	42,566,298	42,173,026
Joint Wald statistic for nonlinear terms	189.93	1,024.66
Nonlinear degrees of freedom	2	2
p-value	<0.001	<0.001
Approximate minimum	0.0239	0.0624
Effect at minimum	−2.479	−3.723
95% CI at minimum	[−2.526, −2.431]	[−3.749, −3.696]
Approximate zero-crossing point	0.112	0.503
Approximate point at which 95% CI becomes positive	0.116	0.514
Share of observations beyond zero crossing	0.329%	0.263%

Notes: The dependent variable is the disruptiveness percentile. Restricted cubic spline models are estimated separately for the proportions of new words and new phrases. Because both measures contain a substantial mass at zero, each specification includes an indicator distinguishing papers with zero new-concept proportions from papers with positive proportions, while the spline function is estimated over the positive-value distribution. All models include the full set of control variables and publication-year and field fixed effects. The joint Wald statistic tests the null hypothesis that all nonlinear spline terms are jointly equal to zero. Effects at the approximate minima are expressed as adjusted differences in disruptiveness percentile points relative to papers with zero new-concept proportions. The zero-crossing point is the value at which the fitted association first changes from negative to positive, and the confidence-bound crossing point is the value at which the lower bound of the 95% confidence interval first exceeds zero. Approximate minima and crossing points are calculated from a dense prediction grid and should be interpreted as descriptive features of the fitted curves rather than as formally estimated structural thresholds.

5 Discussion

This study examines the association between the introduction of new concepts and scientific disruptiveness using more than 42 million publications published between 1950 and 2024. Across the baseline specifications, papers introducing new words or new phrases are, on average, less disruptive than papers that do not introduce new concepts. This negative association is observed across most scientific fields and becomes consistently negative from approximately the 1980s onward. At the same time, the restricted cubic spline analyses show that the relationship is not adequately described by a simple linear model. Disruptiveness initially declines as the proportion of new concepts increases, but the fitted association subsequently rises and becomes positive relative to papers with no new concepts only in the sparsely populated upper tail of the distribution.

These results reinforce the distinction between conceptual novelty at the time of publication and disruptive influence as reflected in subsequent citation behavior. The introduction of new terminology is an ex ante feature of scientific production, whereas disruptiveness is an ex post measure of how later research relates a focal paper to its predecessors. The two dimensions need not move together. A paper may introduce a new expression, method, or conceptual combination while still being incorporated into an established research trajectory. Conversely, disruptive influence may arise from work that reorganizes or redirects existing concepts without necessarily introducing a large amount of new terminology. The findings therefore suggest that textual novelty and citation-based disruption capture related but distinct aspects of scientific change.

The historical results indicate that this relationship has changed over time. In earlier decades, introducing new concepts was more often positively associated with disruptiveness, whereas the association became predominantly negative in later periods. Several interpretations are possible. The growth and increasing specialization of science may have expanded the production of narrowly defined terminology, making the first appearance of a term less closely connected to a major conceptual break. New expressions may also be introduced to differentiate methods, datasets, applications, or fields within an existing knowledge structure. Under these conditions, conceptual introduction may

increasingly reflect specialization and elaboration rather than replacement of prior research. These explanations are consistent with the observed temporal pattern, but the present analysis does not directly identify which historical or organizational process is responsible for the change.

The nonlinear results provide a more qualified account of the relationship between conceptual concentration and disruptiveness. For both new words and new phrases, formal Wald tests reject a linear specification. Thus, the average negative association does not apply uniformly across the full range of conceptual novelty. The substantive interpretation of this pattern should nevertheless remain cautious. The positive portion of the fitted relationship occurs only among papers with unusually high concentrations of new concepts. The upper-tail estimates are therefore based on a relatively small part of the publication distribution, and the descriptive confidence intervals in the most extreme ranges are wider. The spline results support a non-monotonic functional form and an upper-tail reversal, but they do not imply that there is a universal level of conceptual novelty at which a paper necessarily becomes disruptive.

One possible interpretation is that isolated new terms are often absorbed into existing scientific frameworks, whereas papers built around a much larger concentration of new terminology may be more likely to depart from established conceptual structures. However, the analysis cannot determine whether high conceptual concentration itself contributes to disruptiveness. It may instead be associated with other characteristics, such as entry into an emerging field, the development of a new methodological system, unusual problem selection, or differences in publication and citation practices. The estimated minima and zero-crossing points are model-based summaries of the fitted curves rather than structurally identified breakpoints. Their locations may change with the spline specification, the measurement of concepts, the composition of the sample, and the limited density of observations in the upper tail.

The citation decomposition helps clarify the overall negative association. Papers introducing new concepts receive more disruptive citations as well as more consolidating citations, indicating that they attract greater subsequent attention. However, the increase is larger for consolidating citations. Later work is therefore more likely to cite both the focal paper and its predecessors, rather than treating the focal paper as a replacement for the prior knowledge base. This pattern is consistent with the interpretation that many new concepts become integrated into existing scientific trajectories. It does not, however, establish the cognitive or social processes through which such integration occurs.

The team-diversity results should also be interpreted as evidence of heterogeneity rather than as direct evidence of a mechanism. Institutionally, internationally, age-, and gender-diverse teams are associated with lower disruptiveness and a lower probability of introducing new concepts, and the negative association between new concepts and disruptiveness is stronger among more diverse teams. These patterns are consistent with the possibility that coordination and integration demands influence the form of knowledge produced by heterogeneous collaborations. At the same time, the interaction analyses do not constitute a mediation test and do not show that coordination costs, consensus-building, or knowledge integration causally generate the observed association. Team composition may also be related to field, project scale, funding, problem complexity, and other unobserved characteristics.

The findings have limited but potentially relevant implications for research evaluation. The introduction of new terminology alone should not be treated as equivalent to disruptive scientific contribution. Evaluation systems that reward visible linguistic novelty may capture some forms of conceptual development while missing whether the work subsequently redirects or consolidates existing knowledge. At the same time, the upper-tail results do not justify using the proportion of new concepts as a direct indicator of transformative potential. High conceptual concentration is rare, difficult to measure precisely, and may reflect several different types of scientific activity. The results therefore support using multiple complementary indicators when assessing novelty and scientific influence rather than privileging either terminology-based measures or citation-network measures in isolation.

Several limitations should be acknowledged. First, the measures of new concepts are based on the first observed appearance of words and noun phrases in publication titles and abstracts. These measures may capture genuine conceptual innovation, but they may also reflect stylistic variation, changes in terminology, differences in indexing, or narrow specialization. They do not directly measure the theoretical importance, semantic distinctiveness, or scientific validity of the concepts introduced.

Second, abstract coverage in OpenAlex is incomplete and varies across periods and fields. Publications without abstracts are represented primarily by their titles, which contain substantially less conceptual information. This limitation may reduce the precision of the measures and may be particularly relevant to the historical analysis, where metadata coverage is less complete. Improvements in archival text coverage or access to full texts could change the identification and timing of first concept appearances.

Third, the CD index captures one specific form of citation-based disruptiveness. It depends on the observed citation network, reference practices, citation windows, and the assumption that references adequately represent a paper's knowledge base. Research with substantial theoretical, technological, clinical, or societal influence may not receive a high disruption score. The results should therefore not be interpreted as measuring all forms of transformative impact.

Fourth, the observational design does not support causal interpretation. The regressions identify conditional associations after controlling for observed author, team, publication, field, and year characteristics, but unobserved differences may remain. The analyses do not establish that introducing new concepts lowers disruptiveness, that increasing conceptual concentration produces an upper-tail reversal, or that team diversity changes the effect of conceptual novelty. Selection into research topics, collaboration structures, journals, and emerging fields may contribute to the estimated relationships.

Fifth, although restricted cubic splines provide a formal test of nonlinearity, they do not identify a unique structural threshold. The estimated turning regions are sensitive to the locations of knots, the prediction range, the treatment of zero values, and the small number of observations at very high proportions. The nonlinear findings are therefore best understood as evidence that the average association changes across the distribution, rather than as evidence of a fixed transition point.

Future research could examine the semantic content and diffusion of new concepts more directly. Contextual language models may help distinguish terminological variation from substantive conceptual change and identify whether new phrases represent methods, theories, empirical phenomena, or applications. Longitudinal analyses could trace how concepts spread across fields and whether different adoption patterns are associated with disruptive or consolidating citation structures. More direct tests of organizational explanations could combine publication data with information on project formation, funding, team coordination, or changes in collaboration structure. Related analyses of patents, technical standards, and other forms of knowledge production could also clarify whether the nonlinear pattern is specific to scientific publications or appears in other innovation settings.

References

1. Di Bona G, Bellina A, De Marzo G, et al. The dynamics of higher-order novelties. *Nature Communications* 2025; 16: 393. DOI: 10.1038/s41467-024-55115-y.
2. Leahey E, Lee J and Funk RJ. What Types of Novelty Are Most Disruptive? *Am Sociol Rev* 2023; 88: 562–597. DOI: 10.1177/00031224231168074.
3. Arthur WB. The structure of invention. *Research Policy* 2007; 36: 274–287. DOI: 10.1016/j.respol.2006.11.005.
4. Azoulay P, Graff-Zivin J, Uzzi B, et al. Toward a more scientific science. *Science* 2018; 361: 1194–1197. DOI: 10.1126/science.aav2484.
5. Azoulay P, Graff Zivin JS and Manso G. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics* 2011; 42: 527–554. DOI: 10.1111/j.1756-2171.2011.00140.x.
6. Fortunato S, Bergstrom CT, Boerner K, et al. Science of science. *Science* 2018; 359. DOI: 10.1126/science.aao0185.
7. Ma B and Kang L. Knowledge structure and innovation performance in innovation communities. *J Inf Sci* 2025: 01655515251359750. DOI: 10.1177/01655515251359750.
8. Wu LF, Wang DS and Evans JA. Large teams develop and small teams disrupt science and technology. *Nature* 2019; 566: 378–+. DOI: 10.1038/s41586-019-0941-9.
9. Lin Y, Frey CB and Wu L. Remote collaboration fuses fewer breakthrough ideas. *Nature* 2023; 623: 987–991. DOI: 10.1038/s41586-023-06767-1.
10. Park M, Leahey E and Funk RJ. Papers and patents are becoming less disruptive over time. *Nature* 2023; 613: 138–144. DOI: 10.1038/s41586-022-05543-x.
11. Funk RJ and Owen-Smith J. A dynamic network measure of technological change. *Manage Sci* 2017; 63: 791–817. DOI: 10.1287/mnsc.2015.2366.
12. Bloom N, Jones CI, Van Reenen J, et al. Are Ideas Getting Harder to Find? *American Economic Review* 2020; 110: 1104–1144. DOI: 10.1257/aer.20180338.
13. Higham K, de Rassenfosse G and Jaffe AB. Patent Quality: Towards a Systematic Framework for Analysis and Measurement. *Research Policy* 2021; 50: 104215. DOI: 10.1016/j.respol.2021.104215.
14. Hall BH and Jaffe AB. Measuring Science, Technology, and Innovation: A Review. *Annals of Science and Technology Policy* 2018; 2: 1–74. DOI: 10.1561/110.00000005.
15. Kuhn TS. Historical structure of scientific discovery. *Science* 1962; 136: 760–764.
16. Wang D, Song C and Barabási A-L. Quantifying Long-Term Scientific Impact. *Science* 2013; 342: 127–132. DOI: 10.1126/science.1237825.
17. Dorsch I. Becoming Metric-Wise: A Bibliometric Guide for Researchers. *Journal of the Association for Information Science and Technology* 2019; 70: 530–532. DOI: 10.1002/asi.24108.
18. Leibel C and Bornmann L. What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics* 2024; 129: 601–639. DOI: 10.1007/s11192-023-04873-5.
19. Liu X, Bu Y, Li M, et al. Monodisciplinary collaboration disrupts science more than multidisciplinary collaboration. *Journal of the Association for Information Science and Technology* 2024; 75: 59–78. DOI: 10.1002/asi.24840.
20. Yang AJ, Yan X, Hu H, et al. Are disruptive papers more likely to impact technology and society? *Journal of the Association for Information Science and Technology* 2025; 76: 563–579. DOI: 10.1002/asi.24947.

21. Wang S, Ma Y, Mao J, et al. Quantifying scientific breakthroughs by a novel disruption indicator based on knowledge entities. *Journal of the Association for Information Science and Technology* 2023; 74: 150–167. DOI: 10.1002/asi.24719.
22. Liang Z, Mao J and Li G. Bias against scientific novelty: A prepublication perspective. *Journal of the Association for Information Science and Technology* 2023; 74: 99–114. DOI: 10.1002/asi.24725.
23. Ruan X, Ao W, Lyu D, et al. Effect of the topic-combination novelty on the disruption and impact of scientific articles: Evidence from PubMed. *J Inf Sci* 2023: 01655515231161133. DOI: 10.1177/01655515231161133.
24. Dirk L. A measure of originality: The elements of science. *Social Studies of Science* 1999; 29: 765–776.
25. Boudreau KJ, Guinan EC, Lakhani KR, et al. Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. *Manage Sci* 2016; 62: 2765–2783. Article. DOI: 10.1287/mnsc.2015.2285.
26. Foster JG, Rzhetsky A and Evans JA. Tradition and Innovation in Scientists' Research Strategies. *Am Sociol Rev* 2015; 80: 875–908. DOI: 10.1177/0003122415601618.
27. Merton RK. The normative structure of science. *The sociology of science: Theoretical and empirical investigations* 1979: 267–278.
28. Arts S, Melluso N and Veugelers R. Beyond Citations: Measuring Novel Scientific Ideas and their Impact in Publication Text. *The Review of Economics and Statistics* 2025: 1–33. DOI: 10.1162/rest_a_01561.
29. Fleming L. Recombinant uncertainty in technological search. *Manage Sci* 2001; 47: 117–132. DOI: 10.1287/mnsc.47.1.117.10671.
30. Hofstra B, Kulkarni VV, Munoz-Najar Galvez S, et al. The Diversity-Innovation Paradox in Science. *Proceedings of the National Academy of Sciences* 2020; 117: 9284–9291. 2020/04/16. DOI: 10.1073/pnas.1915378117.
31. Min C, Chen QY, Yan EJ, et al. Citation cascade and the evolution of topic relevance. *Journal of the Association for Information Science and Technology* 2021; 72: 110–127. DOI: 10.1002/asi.24370.
32. Xiang S, Jiang X, Liu B, et al. SciConNav: Knowledge navigation through contextual learning of extensive scientific research trajectories. *Journal of the Association for Information Science and Technology* 2025. DOI: 10.1002/asi.25005.
33. Priem J, Piwowar HA and Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv preprint* 2022.
34. Van Buskirk I, Clauset A and Larremore DB. An Open-Source Cultural Consensus Approach to Name-Based Gender Classification. *Proceedings of the International AAAI Conference on Web and Social Media* 2023; 17: 866–877. DOI: 10.1609/icwsm.v17i1.22195.
35. Jones BF, Wuchty S and Uzzi B. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science* 2008; 322: 1259–1262. DOI: 10.1126/science.1158357.
36. Wagner CS, Whetsell TA and Mukherjee S. International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy* 2019; 48: 1260–1270. DOI: 10.1016/j.respol.2019.01.002.
37. Yang AJ, Xu H, Ding Y, et al. Unveiling the dynamics of team age structure and its impact on scientific innovation. *Scientometrics* 2024; 129: 6127–6148. DOI: 10.1007/s11192-024-04987-4.
38. Yang Y, Tian TY, Woodruff TK, et al. Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences* 2022; 119: e2200841119. DOI: 10.1073/pnas.2200841119.
39. Kedrick K, Levitskaya E and Funk RJ. Conceptual structure and the growth of scientific knowledge. *Nature Human Behaviour* 2024; 8: 1915–1923. DOI: 10.1038/s41562-024-01957-x.
40. Xu F, Wu L and Evans J. Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences* 2022; 119: e2200927119. DOI: 10.1073/pnas.2200927119.