Check for updates

Received: 21 December 2023 | Revised: 10 May 2024 | Accepted: 18 July 2024

DOI: 10.1002/asi.24947

RESEARCH ARTICLE

JASIST WILEY

Are disruptive papers more likely to impact technology and society?

Alex J. Yang[1,2] | **Xiaohui Yan**[1] | **Haotian Hu**[1,3] | **Hanlin Hu**[1] | **Jia Kong**[1,4] | **Sanhong Deng**[1,5]

[1]School of Information Management, Nanjing University, Nanjing, China

[2]National Experiment Base for Intelligent Evaluation and Governance, Fudan University, Shanghai, China

[3]Jiangsu Academy of Agricultural Sciences, Nanjing, China

[4]School of Information, Guizhou University of Finance and Economics, Guiyang, China

[5]Key Laboratory of Data Engineering and Knowledge Services in Provincial Universities, Nanjing University, Nanjing, China

Correspondence
Alex J. Yang and Sanhong Deng, School of Information Management, Nanjing University, Nanjing 210023, China.
Email: alexjieyang@outlook.com (A.J.Y.) and sanhong@nju.edu.cn (S.D.)

Funding information
Open Fund for Innovative Evaluation, Grant/Award Number: #CXPJ2024004

Abstract

In exploring the intersection of scholarly research with technological advancement and societal impact, our analysis delves into nearly 40 million research papers spanning from 1950 to 2020 across all fields of study in science. Our scrutiny reveals an intriguing phenomenon: papers characterized by a higher CD index, often considered transformative, paradoxically exhibit a diminished propensity to influence technological and societal domains. This observation suggests a latent bias against the CD index, prompting a deeper inquiry into its implications. To unravel this trend, we introduce the concept of "disruptive citation," a nuanced metric gauging the absolute disruptive impact of papers. Notably, papers drawing higher disruptive citations exhibit a significantly higher probability to influence both technological and societal spheres. Upon examining the heterogeneity across years and fields, we identify a bias against the CD index predominantly in the last two decades and within STEM fields. However, the positive effects of disruptive impact remain consistent across all years and fields. Our findings remain robust even when employing alternative measures of disruptive impact and controlling for total citations. By shedding light on these dynamics, our study seeks to enrich discussions regarding the recognition and role of disruptive scientific endeavors in shaping our world.

1 | INTRODUCTION

Scientific inquiry has perennially stood as a beacon of progress, driving transformative shifts in knowledge and reshaping societal landscapes (Azoulay et al., 2018; Fortunato et al., 2018). However, the realm of scientific discovery often teeters between the realms of incremental progress and disruptive innovation, encapsulating the dichotomy between established paradigms and groundbreaking discoveries (Bower, 1995; Funk & Owen-Smith, 2017; Lin, Frey, & Wu, 2023; Park et al., 2023; Wu et al., 2019; Xu et al., 2022). While the concept of disruptive science—ushering in novel paradigms and challenging conventional wisdom—has garnered attention, its intricate relationship with technological advancement and societal impact remains an enigma.

The perception of science as a public good coexists with its perceived exclusivity, often detached from immediate public benefit (Azoulay et al., 2019; Petersen et al., 2023; Yin et al., 2021, 2022). The divide between scientists and policymakers, shaped by divergent interests and language barriers, has influenced research utilization in policy-making. However, evolving interaction models highlight a more integrated process, showcasing the co-evolution of science and practical applications in the era of big data. Our investigation focuses on four categories of science linked to technology and society: papers cited in patented inventions, used in clinical trials, featured in

 © 2024 Association for Information Science and Technology.

notable news outlets, and disseminated through social media (see Figure 1). Our findings contribute to understanding science's public utility, offering insights to researchers and policymakers committed to fostering impactful scientific endeavors.

The CD index, introduced by Funk and Owen-Smith (2017), meticulously assesses the disruptive nature of papers by scrutinizing structural features within the deep citation network. However, while the CD index provides insight into the proportion of disruptive impact relative to total impact, it exhibits limitations and biases in various contexts (Leibel & Bornmann, 2023; Yang et al., 2024). In this study, we further propose a new metric termed the "disruptive citation," intended to complement the existing CD index, which measures the relative disruptive potential of academic papers. The disruptive citation seeks to assess the absolute disruptive impact of papers, thereby addressing some of these limitations. Central to our investigation is the differentiation between the CD index and disruptive impact, with the latter encompassing both citation impact and disruption levels (Lin et al., 2022; Yang, Hu, et al., 2023). This conceptual distinction forms the basis of our exploration into the relationship between scientific disruption and the probability of papers to be linked to technology and society.

Our study endeavors to shed light on the interplay between the disruptive nature of scientific papers and their impact on technology and society. To achieve this objective, we utilize an extensive repository comprising nearly 40 million research papers sourced from the Microsoft Academic Graph (MAG) dataset, spanning diverse academic disciplines and journals from the years 1950 to 2020. By integrating multiple datasets, we conduct an in-depth analysis of the correlation between scientific disruption and the probability of papers being associated with the realms of technology and society.

The subsequent sections of this paper are organized as follows: The Theoretical Background section establishes foundational concepts related to scientific disruption and its role in public-use science, laying the theoretical groundwork for our hypotheses. The Data section details the data collection process. The Variables and Models section outlines the variables and empirical research design. The Results section presents the findings and heterogeneity analysis. Finally, the Discussions section provides implications, limitations, and avenues for future research.

2 | THEORETICAL BACKGROUND AND HYPOTHESES

2.1 | Science of public benefit

Traditionally, the domain of science has been perceived as an exclusive enclave, confined within an ivory tower inaccessible to the broader public, often appearing disconnected from direct societal benefits (Li et al., 2017; Yin et al., 2022). While theoretically acknowledged as a public good (Stiglitz, 1999), its intricate and specialized nature impedes widespread public recognition. The "two-communities theory" (Caplan, 1979) delineates scientists and policymakers into distinct spheres marked by limited common language and divergent interests. This division is accentuated by supply-side and demand-side models, accentuating either knowledge production by scientists or problem-solving efforts by policymakers (Weiss, 1979), thereby shaping the utilization of research in policy

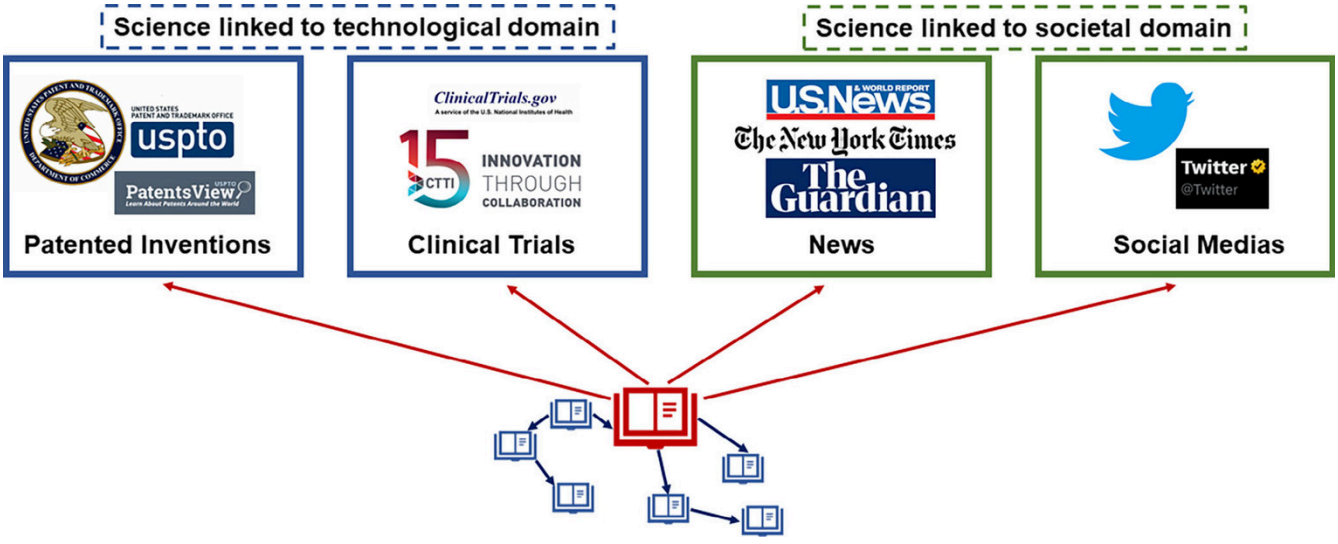


FIGURE 1 Science linked to the technological and societal domains.

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

formulation. Recent interaction models have aimed at portraying a more integrated process, illustrating the co-evolution of science and policy (Yin et al., 2021). In the era of big data science, the availability of diverse datasets, such as patent citations (Ahmadpoor & Jones, 2017), clinical trials (Azoulay et al., 2019), news sources (Yin et al., 2022), and social media metrics (Costas et al., 2021; Sugimoto et al., 2017), has allowed certain scientific endeavors to transcend disciplinary boundaries. These endeavors are cited beyond scientific circles, indicating a tangible link between scientific research, technological advancements, and public discourse.

These links serve multifaceted purposes, encompassing instrumental uses where knowledge directly tackles real-world issues (Jasanoff, 1987) and conceptual uses where research shapes policymakers' perspectives. Beyond-science citations tactically support or challenge specific ideas, showcasing the intricate semantic nuances of policy-science citations. However, grasping these uses on a large scale remains a significant challenge (Yin et al., 2022). The advent of large-scale open-access datasets has, nevertheless, facilitated a deeper understanding of the characteristics underlying science linked to public utilization.

2.2 | Bias against the CD index

Disruptive innovation encapsulates pioneering discoveries or methodologies that confront existing paradigms, marking a departure from incremental progress toward radical shifts in knowledge frameworks (Kuhn, 1962; Merton, 1973). The quantification of disruption in science involves an analysis of citation links within the knowledge network, encompassing both forward and backward impacts of papers, contributing to an extensive knowledge flow network. The CD index, introduced by Funk and Owen-Smith (2017), meticulously assesses the disruptive nature of technologies by scrutinizing structural features within the deep citation network. Building upon this index, Wu et al. (2019), Park et al. (2023), and Lin, Frey, and Wu (2023) have extended its application to scientific literature, furnishing a quantitative measure for evaluating the disruptive nature of scientific knowledge. Leibel and Bornmann (2023) offered a comprehensive review of the CD index, delineating its rationale, capabilities, extensions, and limitations.

While the concept of disruption stands as a pivotal element in scientific breakthroughs and paradigm shifts, it is noteworthy that papers with higher CD indices often do not inherently exhibit higher scientific impact. Conventional, more efficient teams or novel knowledge combinations do not consistently instigate disruptive shifts in science. This observation collectively signifies a prevailing bias against the CD index within the scientific community (Wang et al., 2017; Wuestman et al., 2020). Liang et al. (2023) observed a U-shaped relationship between the CD index and 3- and 5-year citations, indicating that highly disruptive papers do not necessarily yield higher citation impact.

Moreover, this bias against the CD index extends to team dynamics. Wu et al. (2019) identified that smaller teams disrupt science more effectively than larger teams, challenging the presumption that larger teams foster disruptive innovation. Similarly, Xu et al. (2022) found that flat team structures, as opposed to centralized ones, are more conducive to disruptive scientific endeavors. Lin, Frey, and Wu (2023) discovered that remote teams exhibit a lower propensity for disruptive science compared to on-site teams, while Liu et al. (2024) highlighted that monodisciplinary collaboration disrupts science more effectively than multidisciplinary collaboration. These findings indicate that certain team characteristics, often associated with fostering novelty and impact, paradoxically contribute less to disruptive innovation.

Although there exists a dearth of studies examining whether this bias against the CD index extends to public recognition areas such as technological and societal impact, we infer from existing research that such a bias likely exists. Consequently, we propose:

> **H1.** Papers with higher CD index are less likely to influence technology and society.

2.3 | Disruptive impact

In delineating the landscape of scholarly impact, the CD index serves as a metric indicative of the relative proportion of disruptive citations, whereas the concept of disruptive impact elucidates the absolute disruptive citations—those singularly referencing the focal paper (FP) without citing its antecedents (Lin et al., 2022; Yang, Hu, et al., 2023). Implicit within the measure of disruptive impact is the amalgamation of both citation impact and disruption levels, affording precedence to broader citations.

While the disruptive citation is a parameter of the CD index, it represents a distinct metric due to its ability to mitigate various biases. Notably, unlike the CD index which tends toward centralization at 0, the disruptive citation adheres to a power-law distribution (see Section 4.1) akin to other network indices such as citations, degree, and PageRank (Zeng et al., 2017). Furthermore, the disruptive citation metric disregards the influence of the FP's references, thereby avoiding bias stemming from the citing behavior of said paper (Ruan et al., 2021; Yang et al., 2024). Moreover, its expansive range facilitates a nuanced differentiation between papers with high or low disruptive impact, thereby

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enhancing its efficacy in ranking scholarly contributions. Recent investigations indicate that disruptive citations demonstrate superior efficacy in identifying milestone papers and even laureates (Lin et al., 2022; Yang, Hu, et al., 2023), a capability not shared by the CD index.

As such, we posit that the concept of disruptive impact constitutes a form of impact inherent to scholarly papers (Lin et al., 2022; Yang, Hu, et al., 2023). Notably, technological and societal impacts represent alternative facets of paper impact. We contend that each of these dimensions encapsulates broader aspects of the value and recognition accorded to scientific research, extending beyond mere citation impact. At its core, the dissemination of scientific knowledge into the technological and societal domain often entails the confrontation of real-world challenges and societal quandaries (Azoulay et al., 2019). This endeavor necessitates a departure from conventional methodologies, fostering innovative thinking and the exploration of novel scientific avenues (Shi & Evans, 2023). Moreover, such endeavors are frequently impelled by the exigency of practical applications and palpable societal dividends (Yin et al., 2022). The societal reverberations of such research endeavors compel scientists toward the unexplored, engendering experimentation with innovative methodologies and the relentless expansion of the boundaries of existing knowledge to tackle exigent societal dilemmas (Fortunato et al., 2018). The intrinsic impetus to effectuate meaningful change and enhance societal welfare substantiates the contention that research conducted for public good inherently engenders heightened levels of disruptive impact. Bower and Christensen (1995a, 1995b) assert that disruptive innovation can significantly influence both technology and society. Importantly, scientific endeavors intersecting with the technological and societal domains are poised to yield profound technological and societal impacts, thereby aligning with heightened disruptive impact.

Thus, we hypothesize a positive interrelationship among these distinct impacts, collectively enriching our comprehension of the multifaceted contributions of scientific scholarship.

> **H2.** Papers with higher disruptive impact are more likely to influence technology and society.

3 | DATA

3.1 | Paper and citation dataset

This study harnesses the MAG, renowned for its exhaustive repository of scholarly endeavors (Wang et al., 2020). Spanning from 1800 to 2021, this dataset comprises an expansive corpus of over 200 million documents, encompassing journal articles, conference proceedings, preprints, and assorted research publications. MAG provides a meticulously curated framework for the delineation of research fields, facilitated by cutting-edge machine learning algorithms. At its primary level, 19 overarching research domains are demarcated, further distilled into 292 subfields at the secondary level.

Specifically targeting research articles published between 1950 and 2020, our inquiry concentrates solely on journal articles garnering at least one citation and featuring at least one reference—an essential criterion for the computation of the CD index (Wu et al., 2019). Other document types are deliberately omitted due to potential variations in citation patterns. Through this way, we get 39,643,629 journal research articles.

While the citation timeframe (Sinatra et al., 2016) critically informs the computation of the CD index and the identification of disruptive citations, we abstain from imposing constraints on the citation timeframe within the primary discourse of our study. This methodological stance resonates with recent studies published in Nature (Lin, Frey, & Wu, 2023; Wu et al., 2019). Moreover, supplementary analyses employing 5-year windows (1950–2017) and 10-year windows (1950–2011) are presented in the Supporting Information, offering additional insights.

3.2 | Science linked to technology and society

Understanding the intricate interplay between science, technology, and society is of paramount importance. However, the task of comprehensively unraveling this nexus is formidable due to the scarcity of pertinent data (Yin et al., 2022). It is only in recent years that we have gained access to an abundance of publicly available large-scale datasets. Drawing inspiration from the seminal work of Azoulay et al. (2019), who assert that the predominant technological impact of scientific research manifests through patented inventions and subsequent clinical trials, we embark on a parallel trajectory of inquiry. In this pursuit, we extend our purview to encompass the societal impact, capitalizing on data streams from Twitter and News media, both of which are available to link to the MAG dataset and offer high-quality data.

In this study, we focus on four distinct categories of science that gain recognition within the technological and societal domain. The first category revolves around scientific research garnering citations from patented inventions. The translation of scientific knowledge into practical applications within the marketplace has been a focal point in the domains of science and innovation literature (Azoulay et al., 2019; Fleming et al., 2019). Robust datasets concerning patenting activities, including extensive metadata on

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

patent assignees, inventors, legal aspects, internal citations, and full-text information, are obtainable from reputable sources like the US Patent and Trademark Office, PatentsView, and the European Patent Office. Paper-patent citations serve as a conduit for the transfer of knowledge from science to technology (Azoulay et al., 2019). Leveraging recent advancements in linking scientific papers to patents, Marx and Fuegi (2022) compiled an extensive dataset capturing citations from patents to papers, which we employ to examine the recognition of science by patented inventions.

The second category delves into scientific research integrated into clinical trials (Azoulay et al., 2019; Li et al., 2017). In biomedical research, comprehending the translation of knowledge from laboratory settings to clinical applications holds paramount importance. Publicly available clinical study records, generously furnished by platforms like ClinicalTrials.gov, provide comprehensive data compiling information on drugs, regulations, publications, and other pertinent facets (Packalen & Bhattacharya, 2020). The clinical linkage data, thoughtfully curated within the SciSciNet dataset (Lin, Yin, et al., 2023), constitute the cornerstone of our inquiry into the integration of scientific research within clinical trial frameworks.

The third and fourth categories encompass scientific research prominently featured in news outlets and disseminated through social media, respectively. Investigating how science is portrayed in the media stands as a critical area of inquiry within the science of science community (Kreps & Kriner, 2020). Newsfeed mentioned in Crossref (Hendricks et al., 2020) establishes connections between papers and diverse news sources, facilitating access to the latest references to scientific research. Simultaneously, Twitter mentions in Crossref establish connections between science and Twitter users, providing an avenue for investigating discussions surrounding science on social media platforms.

It is imperative to acknowledge the existence of potential missing values within our dataset, attributable to inherent limitations within the data sources employed. Nevertheless, the confluence of these disparate datasets engenders invaluable insights into the expansive purview and profound impact of scientific research as it converges with technology and societal dynamics.

4 | VARIABLES AND MODELS

4.1 | The CD index and disruptive impact

The CD index (Funk & Owen-Smith, 2017; Park et al., 2023; Wu et al., 2019) serves as a quantitative measure to assess the disruptive potential of scientific papers. This index is defined by the equation:

$$\text{CD index} = p_D - p_C = \frac{n_{DC} - n_{CC}}{n_{DC} + n_{CC} + n_{RC}} \quad (1)$$

where n_{DC} (the number of disruptive citations) represents the count of publications citing the FP without referencing its cited sources, n_{CC} (the number of consolidating citations) denotes the count of publications citing FP and its references, and n_{RC} represents the count of publications citing FP's references without citing FP itself.

As depicted in Figure 2b,c, a lower CD index indicates extensive citation of references by papers referencing the FP, aligning with established knowledge and consolidating existing paradigms. Conversely, higher CD index signify highly disruptive papers capable of inducing paradigm shifts. Understanding disruption sheds light on knowledge dissemination patterns and scientific paradigm shifts (Yang, Deng, et al., 2023; Yang, Wang, et al., 2023). Given the highly skewed distribution of the CD index and its non-count nature, our approach involves transforming it into percentile variables. This transformation entails ranking papers based on their CD index, ranging from the highest CD index (assigned the percentile value of 1) to the lowest CD index (assigned the percentile value of 0). This conversion allows for a standardized representation of the CD index across the dataset, spanning from 0 to 1.

In addition to the CD index indicating the relative proportion of disruptive impact, we utilize disruptive citation to measure the absolute disruptive impact of papers:

$$\text{Disruptive citation} = n_{DC} \quad (2)$$

As shown in Figure 2d,e, disruptive citations exhibit power law distributions similar to other network-based metrics, providing a nuanced understanding of the disruptive impact of papers. Furthermore, the disruptive citation metric disregards the influence of the FP's references, thereby avoiding bias stemming from the citing behavior of said paper (Ruan et al., 2021; Yang et al., 2024). Moreover, its expansive range facilitates a nuanced differentiation between papers with high or low disruptive impact, thereby enhancing its efficacy in ranking scholarly contributions. Disruptive citation proves effective in identifying milestone papers and laureates (Lin et al., 2022; Yang, Hu, et al., 2023) and offers insights into the relationship between scientific impact and disruption.

4.2 | Control variables

4.2.1 | Team level controls

Recent studies indicate that small and on-site teams may engender the emergence of novel paradigms (Lin,

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

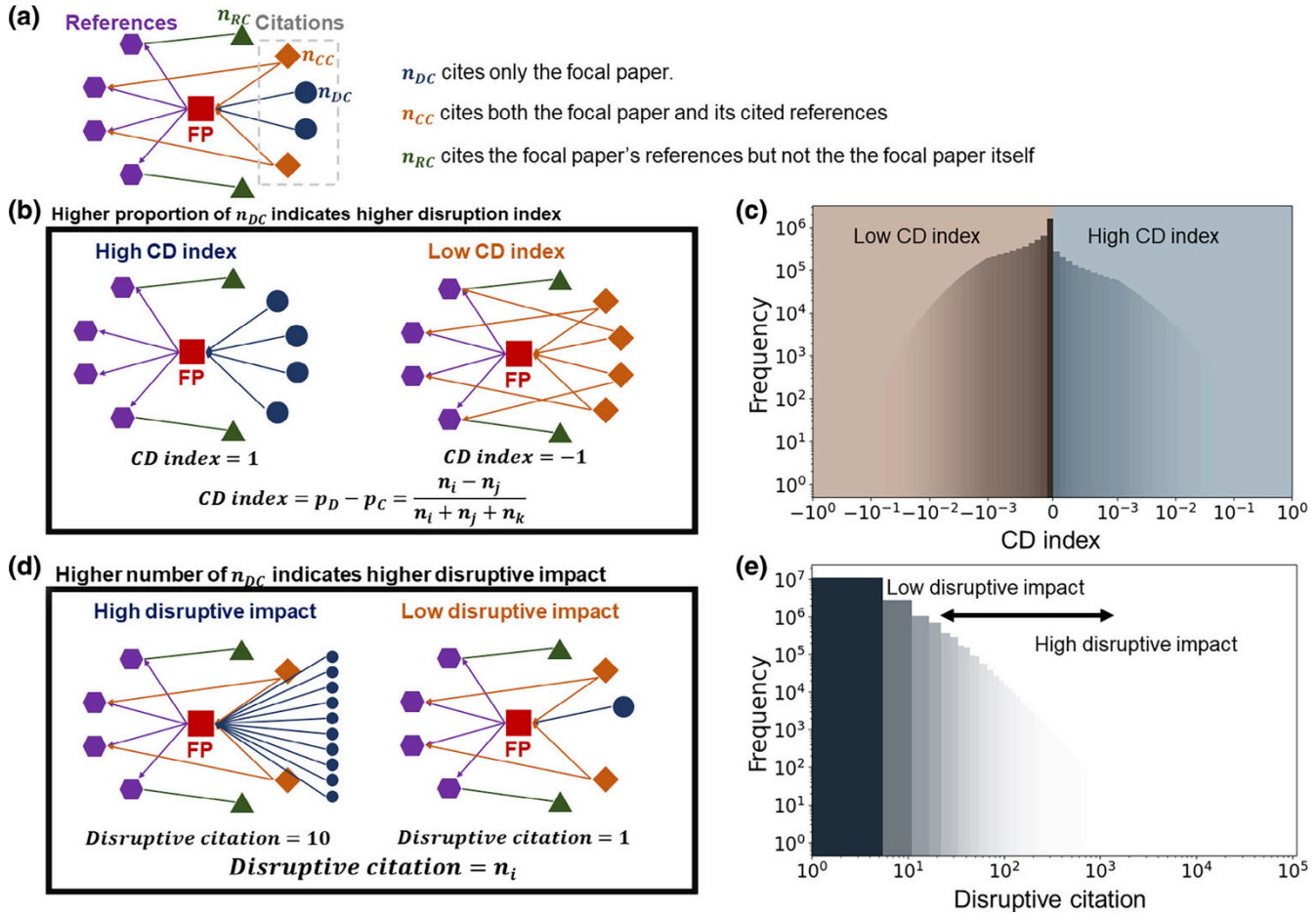


FIGURE 2 Assessing the CD index and disruptive impact in scientific papers.

Frey, & Wu, 2023; Wu et al., 2019). Furthermore, recent research posits that monodisciplinary teams may yield more disruptive outcomes compared to their multidisciplinary counterparts (Liu et al., 2024). To attenuate potential confounding variables, we incorporate the number of authors contributing to a research endeavor (Wuchty et al., 2007). Moreover, we incorporate the dimension of international collaborations, recognizing their global reach and potential impact (Lee et al., 2019). To operationalize and address the complexity of collaborative research, we assess the interdisciplinarity of team members as an indicator of the breadth of accessible knowledge (Liu et al., 2024). This assessment entails delineating the primary field of each scientist using the 19 top-level MAG field-of-study labels. This categorization framework enables the differentiation of monodisciplinary teams, comprising members solely from one discipline, from interdisciplinary teams comprising individuals representing diverse fields. Both the international and interdisciplinary teams are set into dummy variables in our analysis.

4.2.2 | Paper level controls

Grant funding can significantly impact the level of innovation in research papers and their subsequent technological applications (Azoulay et al., 2019). Therefore, we include a variable to ascertain whether a paper received funding from the National Institutes of Health or the National Science Foundation (Yang, 2024). Moreover, we incorporate a control for the number of references cited in each paper. This control is essential as it can influence both the disruption measures and impact of papers.

Interdisciplinarity also plays a crucial role in both the disruptive innovation and impact of scholarly papers (Gates et al., 2019). To measure this multifaceted quality, we adopt the Rao-Stirling index (Stirling, 2007). For each paper i, we commence by mapping its references into a vector $R_i = \left\{r_{i1}^{\text{field}_1}, r_{i2}^{\text{field}_2}, ..., r_{ik}^{\text{field}_k}, ...\right\}$, where the field_k signifies the field of each reference, delineated at the second level within the MAG. Subsequently, we focus on a specific field m and consider all papers published therein during year t. We aggregate the paper vectors within this field to construct the knowledge vector V_{mt} for field m

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and year t. Utilizing the cosine distance metric, we then compute the distance between fields m and n in year t as $d_{mnt} = 1 - CosSim(V_{mt}, V_{nt})$. The Rao-Stirling index is defined as:

$$\text{Rao} - \text{Stirling} = \sum_{m \neq n} d_{mnt} p_m p_n \quad (3)$$

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 adeptly captures the nuanced balance, variety, and disparity present within scholarly works (Stirling, 2007).

4.2.3 | Field, journal, and year fixed effects

To mitigate the inherent heterogeneity across scientific fields, journals, and temporal variations (Chu & Evans, 2021; Park et al., 2023), three-level fixed effects are integrated into the regression models. Firstly, the 292 s-level fields in the MAG dataset are utilized to control for variations among distinct scientific domains. Secondly, journal fixed effects are incorporated to account for the heterogeneity related to publication outlets, encompassing a comprehensive total of 39,893 distinct journals. Lastly, year fixed effects are integrated to capture temporal variations and trends over time.

4.3 | Empirical specification

To evaluate the potential of disruptive research in contributing to technological and societal advancements, evidenced by its association with patents, clinical trials, news coverage, and social media dissemination, a logistic regression model is employed. This methodology facilitates the estimation of the impact of CD index on the probability of scientific papers being cited by these diverse sources. The logistic regression Equation (4) estimates the probability of impacting technology and society:

$$P(Y_i = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1[\text{CDindex}_i] + \beta_2[\text{Controls}] + Y_t + F_f + J_j)}} + \in_i \quad (4)$$

where i represents the article, t the year of publication, f the field, and j the journal. The CD index variables are converted into percentile variables, ranging from 0 to 1, while all counting variables are natural log-transformed within the models. The model structure accounts for numerous factors: Y_t denotes year fixed effects, F_f captures second-level field fixed effects, J_j represents journal fixed effects, and $\in_{itfj}$ denotes the error term.

Furthermore, the disruptive citation serves as the core independent variable to test the effect of disruptive impact on the likelihood of papers being linked to technology and society. Equation (5) estimates the probability:

$$P(Y_i = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1[\text{Disruptive citation}_i] + \beta_2[\text{Controls}] + Y_t + F_f + J_j)}} + \in_i \quad (5)$$

5 | RESULTS

5.1 | Distribution and descriptive analysis

The examination of paper distributions linked to technological and societal domains, encapsulated in Figure 3, unfolds within a vast repository comprising nearly 40 million scientific papers. A notable finding emerges regarding the relative scarcity of papers associated with societal links. Specifically, within this extensive dataset, 9.48% of papers are linked to patents, highlighting a limited intersection between scientific research and patenting activities. Additionally, a mere 0.77% is associated with clinical trials, underlining a relatively modest integration of research into clinical applications. Further, only 0.55% find mention in news sources, while a moderate 6.14% receive recognition through tweets, emphasizing the varying degrees of engagement and visibility of scientific research within different societal spheres (see the subset in Figure 3b).

A deeper temporal examination reveals intriguing trends in the integration of science into public use. Figure 3a portrays an exponential growth in the total paper count over time, reflecting the expansive growth in scientific knowledge production. This increase is mirrored by a proportional rise in papers associated with technology and societal benefit. However, it is noteworthy that although the overarching trend suggests an escalation, there appears to be a marginal downturn in the percentage of papers linked with patents and clinical trials since around 2010. This phenomenon may stem from two plausible factors: firstly, the potential inadequacy in the coverage of recent papers within our dataset; and secondly, the requisite extended duration for papers to gain recognition through technology in patents and clinical trials (Azoulay et al., 2019).

Preliminary analysis involved segregating papers based on their CD index, specifically whether it was

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

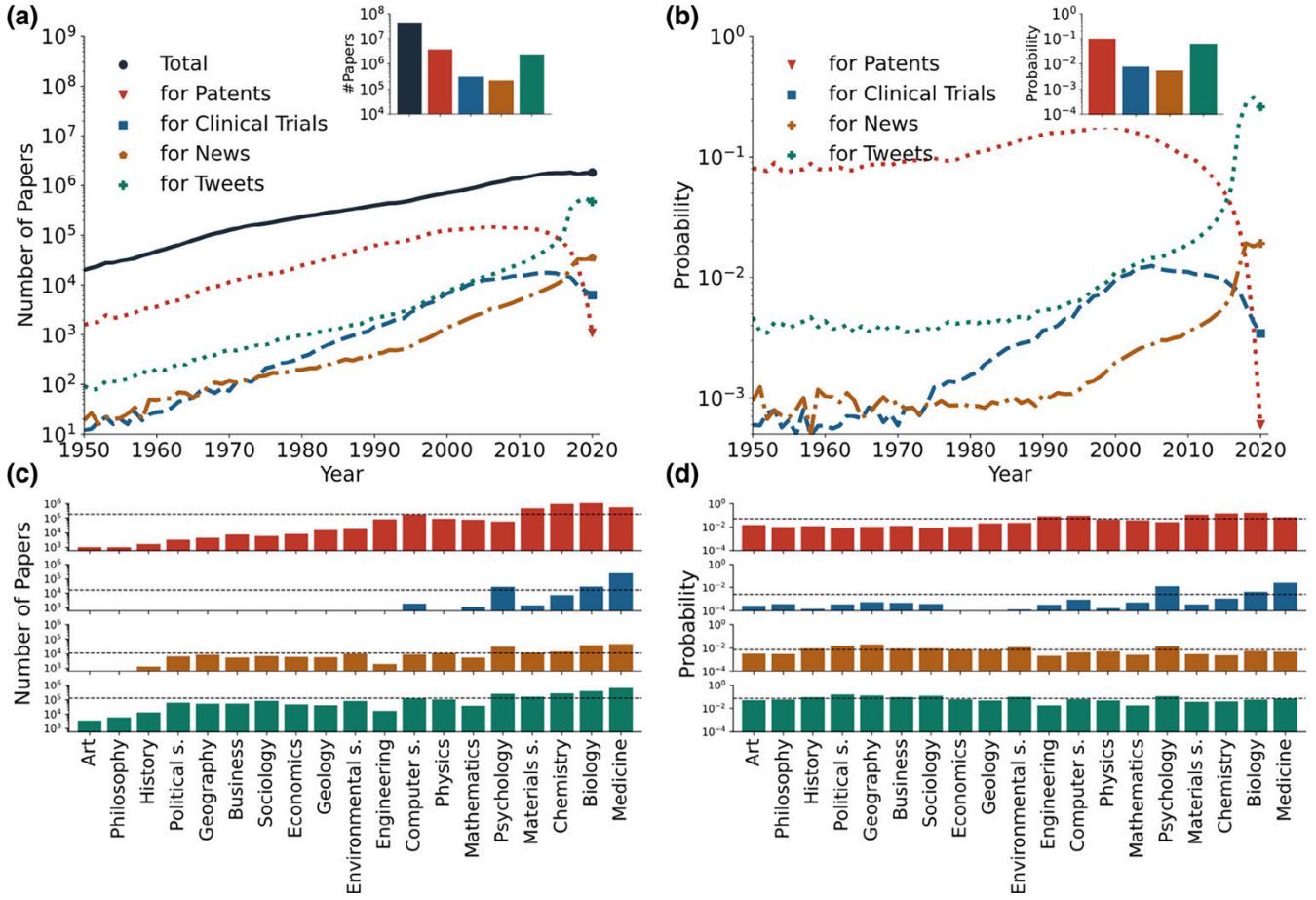


FIGURE 3 Distribution of scientific papers linked to the technological and societal domains.

above or below zero—a positive CD index indicating that FPs receive more disruptive citations than consolidating citations. Remarkably, only 29.3% (11,622,458 out of 39,643,629) of papers in our dataset met this criterion.

Employing a two-tailed t-test to discern differences across variables, as outlined in Table 1, yielded compelling insights. Papers characterized by a CD index above zero not only exhibit a reduced likelihood of being linked to technological and societal domains but also receive fewer citations. Additionally, these papers tend to attract diminished funding, engage smaller and less diverse teams, and possess less diverse knowledge bases. These findings align with and corroborate prior studies (Liu et al., 2024; Park et al., 2023; Wu et al., 2019; Xu et al., 2022), underscoring a discernible bias against the CD index.

5.2 | Effect of the CD index on the likelihood to impact technology and society

Table 2 and Table 3 reveal insights into the relationship between the CD index and the propensity of scientific papers to intersect with technological and societal domains.

Initially employing logit regressions with field and year fixed effects exclusively in models (1) and (5), we discern a persistent inverse association between the CD index and the likelihood of papers being linked to patents or clinical trials. The result suggests that the odds ratio of the most disruptive papers being cited by patents relative to the most consolidative papers is 0.65 (exp(−0.43)), and the odds ratio of the most disruptive papers being associated with clinical trials vis-à-vis the most consolidative papers is 0.27 (exp(−1.33)).

Successive models incrementally incorporate controls, encompassing journal fixed effects, team-level parameters, and paper-level factors. Notably, the introduction of these control variables attenuates the magnitude of the CD index's effect size. This observation underscores the pivotal roles assumed by assorted control variables in shaping biases against disruptive research concerning its technological ramifications. Nevertheless, these augmented models exhibit a conspicuous pattern—a steadfastly negative coefficient for the CD index. In models (4) and (8), wherein all

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Descriptive analysis and comparison with two-tailed t test.

	CD index >0			CD index ≤0					
Variables	**Obs.**	**Avg.**	**Std.**	**Obs.**	**Avg.**	**Std.**	**Diff.**	**T**	***p***
Citation count	11,622,458	28.33	218.3	28,021,171	31.82	85.2	−3.49	−72.32	<0.001
P (Patent)	11,622,458	9.20%	28.9%	28,021,171	9.60%	29.5%	−0.39%	−38.64	<0.001
P (Clinical trials)	11,622,458	0.52%	7.2%	28,021,171	0.88%	9.3%	−0.36%	−117.01	<0.001
P (Newsfeed)	11,622,458	0.38%	6.2%	28,021,171	0.62%	7.9%	−0.24%	−91.33	<0.001
P (Tweet)	11,622,458	4.02%	19.6%	28,021,171	7.01%	25.5%	−2.99%	−357.99	<0.001
Team size	11,622,458	3.49	7.5	28,021,171	4.55	20.0	−1.06	−175.64	<0.001
International team	11,622,458	0.09	0.3	28,021,171	0.17	0.4	−0.08	−630.88	<0.001
Interdisciplinary team	11,622,458	0.28	0.5	28,021,171	0.33	0.5	−0.05	−303.40	<0.001
Funding	11,622,458	17.01	20.1	28,021,171	33.23	31.9	−16.22	−1607.55	<0.001
#Refercence	9,118,058	0.14	0.1	25,717,463	0.16	0.1	−0.01	−423.31	<0.001
Rao-Stirling	11,622,458	0.03	0.2	28,021,171	0.09	0.3	−0.06	−606.92	<0.001

TABLE 2 Negative relationship between the CD index and the likelihood of papers to be linked to technology.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent variables	**P (Patents)**				**P (Clinical trials)**			
CD index (Pct.)	−0.4293*** (0.0020)	−0.2189*** (0.0022)	−0.1975*** (0.0022)	−0.0571*** (0.0024)	−1.325*** (0.0074)	−1.031*** (0.0080)	−0.8945*** (0.0082)	−0.7555*** (0.0094)
ln (Team size)			0.2348*** (0.0012)	0.2148*** (0.0013)			0.4915*** (0.0033)	0.4085*** (0.0034)
International team			−0.0478*** (0.0019)	−0.0711*** (0.0019)			0.0799*** (0.0052)	0.0339*** (0.0054)
Interdisciplinary team			0.1166*** (0.0014)	0.1120*** (0.0015)			−0.0938*** (0.0047)	−0.1436*** (0.0047)
Funding				0.2766*** (0.0020)				0.3970*** (0.0053)
ln (#Reference)				0.4180*** (0.0012)				0.5645*** (0.0032)
Rao-Stirling				−1.605*** (0.0099)				−0.5032*** (0.0344)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs.	39,643,629	38,887,040	38,887,040	34,153,561	39,643,629	20,015,133	20,015,133	18,205,154
Pseudo R^2	0.1328	0.2362	0.23873	0.2398	0.1786	0.1685	0.17701	0.1879

Note: Robust standard errors are reported in parentheses. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

variables are controlled for, the odds ratio of the most disruptive papers being cited by patents relative to the most consolidative papers is 0.94 (exp(−0.06)), and the odds ratio of the most disruptive papers being associated with clinical trials vis-à-vis the most consolidative papers is 0.47 (exp(−0.76)).

In a similar vein, our analysis delves into the correlation between the CD index and papers associated with news sources and tweets. Across the spectrum of model specifications detailed in Table 3, the CD index consistently exhibits a negative impact on the likelihood of papers being linked to news and tweets.

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Negative relationship between the CD index and the likelihood of papers to be linked to society.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variables	P (News)				P (Tweets)			
CD index (Pct.)	−1.219*** (0.0104)	−0.6173*** (0.0112)	−0.5665*** (0.0113)	−0.4709*** (0.0136)	−0.7161*** (0.0033)	−0.2908*** (0.0040)	−0.2762*** (0.0040)	−0.2079*** (0.0047)
ln (Team size)			0.1294*** (0.0040)	0.0677*** (0.0043)			0.0169*** (0.0016)	−0.0257*** (0.0017)
International team			0.0781*** (0.0054)	0.0627*** (0.0057)			0.1603*** (0.0020)	0.1445*** (0.0022)
Interdisciplinary team			0.1347*** (0.0052)	0.1276*** (0.0055)			0.0936*** (0.0019)	0.0898*** (0.0021)
Funding				0.2044*** (0.0068)				0.1161*** (0.0030)
ln (#Reference)				0.3375*** (0.0044)				0.3327*** (0.0017)
Rao-Stirling				0.4766*** (0.0558)				0.0640** (0.0232)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs.	39,643,629	34,184,667	34,184,667	30,012,159	39,643,629	37,995,938	37,995,938	33,403,177
Pseudo R^2	0.1053	0.2265	0.2279	0.2352	0.2909	0.42306	0.4237	0.4345

Note: Robust standard errors are reported in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Commencing with logit regressions with field and year fixed effects exclusively in models (1) and (5), and subsequently incorporating a series of control variables, our results echo those unearthed in the examination of technological impact. In models (4) and (8), encompassing all pertinent variables, the logit regression signifies that the odds ratio of the most disruptive papers being linked to news relative to the most consolidative papers is 0.63 (exp(−0.47)), while the odds ratio of the most disruptive papers being associated with tweets as opposed to the most consolidative papers is 0.81 (exp(−0.21)). In essence, our findings underscore a persistent negative correlation between the CD index and the propensity of papers to be linked to both technological and societal domains.

5.3 | Effect of the disruptive citation on the likelihood to impact technology and society

Table 4 consistently reveals a positive effect of disruptive citation on the likelihood of papers being linked to technology. In models (1) and (5), conducting simple logit regressions with field and year fixed effects exclusively, we observe this positive association. The gradual inclusion of controls further substantiates this positive trend of disruptive citation. Particularly noteworthy are the results in models (4) and (8), where all confounding variables are controlled for. In these models, the results indicate that for each additional point increase in disruptive citation of FPs, the odds ratio of being cited by patents increases by 0.80 points (exp(0.59)-1), while the odds ratio of being associated with clinical trials increases by 1.32 points (exp(0.84)-1).

Table 5 presents our findings concerning the effect of disruptive citation on the likelihood of papers being associated with societal domains. Our analysis consistently reveals a positive effect of disruptive citation. Successively incorporating controls in models (1) through (8) unveils a steadfastly positive relationship between disruptive citation and the likelihood of papers being linked to news sources and tweets. Particularly noteworthy are the results in models (4) and (8), where all confounding variables are considered. In these models, the logit regression indicates that for each additional point increase in disruptive citation of FPs, the odds ratio of being linked to news increases by 1.10 points (exp(0.74) − 1), while the odds ratio of being associated with tweets increases by 0.60 points (exp(0.47) − 1).

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Positive relationship between the disruptive citation and the likelihood of papers to be linked to technology.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variables	P (Patents)				P (Clinical trials)			
ln (Disruptive citation)	0.6401*** (0.0005)	0.6004*** (0.0006)	0.5950*** (0.0006)	0.5855*** (0.0006)	0.9448*** (0.0015)	0.8917*** (0.0018)	0.8787*** (0.0019)	0.8434*** (0.0020)
ln (Team size)			0.1576*** (0.0012)	0.1486*** (0.0013)			0.3412*** (0.0034)	0.3230*** (0.0034)
International team			−0.0707*** (0.0019)	−0.0880*** (0.0020)			−0.0011 (0.0054)	−0.0205*** (0.0056)
Interdisciplinary team			0.1065*** (0.0015)	0.1064*** (0.0015)			−0.1050*** (0.0047)	−0.1325*** (0.0048)
Funding				0.2433*** (0.0021)				0.3032*** (0.0055)
ln (#Reference)				0.2793*** (0.0012)				0.2691*** (0.0032)
Rao-Stirling				−1.124*** (0.0101)				−0.1262*** (0.0345)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs.	39,643,629	38,887,040	38,887,040	34,153,561	39,643,629	20,015,133	20,015,133	18,205,154
Pseudo R^2	0.20237	0.28266	0.28389	0.28304	0.27592	0.24771	0.25129	0.25357

Note: Robust standard errors are reported in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

These findings underscore a clear and consistent positive association between disruptive citation and the likelihood of papers being linked to both technological advancements and societal domains.

5.4 | Heterogeneity in years and fields

Our exploration of the heterogeneity in the CD index's effects involved splitting the sample by year and field. In Figure 4a, we offer outcomes when examining yearly samples, controlling for other confounding factors. Remarkably, the adverse impact of the CD index on the association between papers and technology and society is conspicuously pronounced solely within the last two decades. Papers deemed disruptive published prior to 2000 manifest a propensity toward favorable linkage with technology and societal realms, suggesting an absence of prejudice against disruptive innovation in earlier epochs. However, the augmented sample size in contemporary years, relative to historical ones, engenders an overarching negative trend. These consistent findings span across linkages—patents, clinical trials, news, and tweets. A discernible transition point emerges between 2000 and 2010, signifying a transition from a positive to a negative effect concerning bias against the CD index.

In Figure 4b, our examination across 19 distinct fields reveals patterns across disciplinary boundaries. Conducting regression analyses controlling for other confounding factors, we found the negative effect of CD index to be predominantly observed within STEM fields. This trend partly owes to the relatively lower frequency of papers from social science and art disciplines linked to technology and societal contexts. Furthermore, substantial heterogeneity in effects across different technological and societal contexts within specific fields is evident. Remarkably, domains encompassing Psychology, Chemistry, Biology, and Medicine, wherein clinical trials constitute a prominent research avenue, consistently exhibit substantial negative relationship. Conversely, fields such as Art and History appear less affected, with discernible effects being less pronounced. Furthermore, an exploration of the relationship between the CD index and the probability to be linked to news and social media reveals predominantly negative coefficients, only with exceptions in Humanities and Mathematics.

Our investigation extends to examining the heterogeneity of the positive association between disruptive citation and the likelihood to be linked to technology and society. In Figure 5a, we focus on each year's samples, controlling for other confounding factors. The positive relationship between disruptive citation and the

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Positive relationship between the disruptive citation and the likelihood of papers to be linked to society.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variables	P (News)				P (Tweets)			
ln (Disruptive citation)	0.9019*** (0.0019)	0.7452*** (0.0023)	0.7432*** (0.0023)	0.7424*** (0.0025)	0.4769*** (0.0008)	0.4630*** (0.0010)	0.4612*** (0.0010)	0.4682*** (0.0011)
ln (Team size)			−0.0229*** (0.0040)	−0.0467*** (0.0043)			−0.0402*** (0.0016)	−0.0727*** (0.0017)
International team			0.0664*** (0.0055)	0.0570*** (0.0058)			0.1522*** (0.0020)	0.1403*** (0.0022)
Interdisciplinary team			0.1100*** (0.0053)	0.1061*** (0.0056)			0.0779*** (0.0019)	0.0765*** (0.0021)
Funding				0.1542*** (0.0070)				0.0870*** (0.0030)
ln (#Reference)				0.1678*** (0.0041)				0.2389*** (0.0017)
Rao-Stirling				1.072*** (0.0566)				0.4286*** (0.0235)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs.	39,643,629	34,184,667	34,184,667	30,012,159	39,643,629	37,995,938	37,995,938	33,403,177
Pseudo R^2	0.18302	0.27251	0.27274	0.27899	0.30896	0.43687	0.43726	0.44816

Note: Robust standard errors are reported in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

likelihood to be linked to technology and society remains consistently significant across the past seven decades. This signifies that irrespective of a paper's age—whether older or newly published—if it garners more disruptive citations, it exhibits a higher likelihood of being linked to technological and societal domains.

In Figure 5b, we present the relationship between disruptive citation and the likelihood to be linked to technology and society across all 19 fields. These analyses, controlling for confounding variables, consistently reveal a positively significant effect within each field. This implies that across various academic disciplines—from Biology and Physics to Humanities—papers receiving more disruptive citations are more likely to be associated with technology and societal contexts. These findings underscore the consistent and positive association between disruptive impact and the likelihood of papers being linked to technology and society across temporal and disciplinary boundaries.

5.5 | Robustness analysis

We employ alternative disruptive impact metrics to bolster the robustness of our findings. Following the proposition by Funk and Owen-Smith (2017), alongside the CD index, they introduced the mCD index, calculated as the product of the raw CD index and the citation counts. Employing the mCD index as an alternative measure of disruptive impact, we scrutinize its relationship with the propensity to be associated with technology and society. The findings shown in Tables S2 and S3 consistently reveal a significantly positive association. Hence, we affirm H2, positing that papers with greater disruptive impact are more prone to be linked to the domains of technology and society.

Moreover, to account for the fact that disruptive citations represent a portion of total citations, we introduce total citations as control variables in our regression models. The results presented in Tables S4 and S5 reveal a positive association between the disruptive citation and the likelihood of papers to impact technology and society. Particularly noteworthy is the consistently significant and robust nature of the positive coefficients observed in the domains of Patents, News, and Tweets. Even within the domain of clinical trials, where coefficients tend to be negative without full controls, upon accounting for all relevant variables, the coefficient remains positive. These findings underscore a persistent positive correlation between disruptive impact and the probability of papers

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

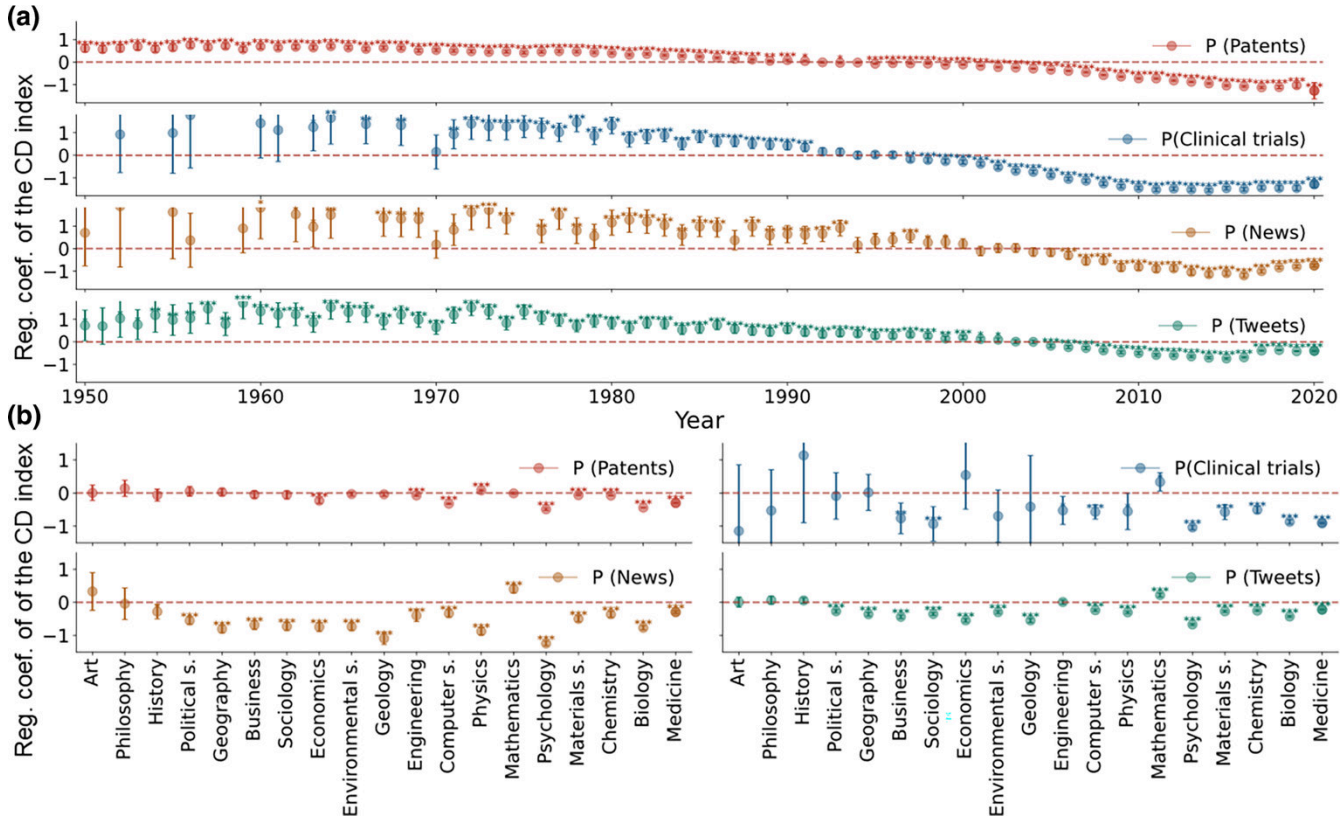


FIGURE 4 Heterogeneity of the relationship between the CD index and the probability of papers being linked to technological and societal domains. Each scatter plot depicts the logit regression coefficient of the CD index on the likelihood of papers being linked to patents, clinical trials, news, or tweets, with error bars representing 95% confidence intervals based on robust standard errors. All control variables have been included; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

being linked to technological and societal domains, a relationship that remains significant even after controlling for overall scientific impact.

We also conduct sensitivity analyses by setting and adjusting the citation windows to assess the robustness of our findings. The results illustrated in Tables S6–S13 indicate that when accounting solely for year and field fixed effects, the 5- or 10-year CD index exhibits a negative relationship with the probability of papers being linked to technology and society, consistent with our primary findings. However, upon inclusion of additional control variables, this relationship turns positive. One plausible explanation for this shift is twofold. Firstly, employing either a 5- or 10-year window for analysis necessitates the exclusion of recent observations. Notably, our analysis underscores that the negative effects of the CD index are consistently observed within the recent 10–20 years. Consequently, this exclusion may skew the average effect toward a positive trajectory. However, it is imperative to acknowledge the potential existence of alternative mechanisms contributing to this outcome. While our study refrains from delving deeper into this matter, it serves as a catalyst for future investigations aimed at unraveling the underlying complexities. Thus, elucidating these intricacies remains a prospective avenue for scholarly inquiry.

Contrarily, even with the adoption of both 5- and 10-year disruptive citation metrics, the affirmative positive relationship between the disruptive impact and the propensity for papers to be associated with technology and society persists uniformly across all domains.

6 | DISCUSSION

6.1 | Implications

This study makes contributions to various aspects of disruptive innovation in science and its technological and societal implications. The stark negative correlation between the CD index and the probability of papers being linked to technological and societal domains paints a compelling picture. It unveils a systemic bias against disruptive research (Liang et al., 2023; Wang et al., 2017), illuminating the hurdles unconventional ideas face in garnering acceptance within the scientific community,

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

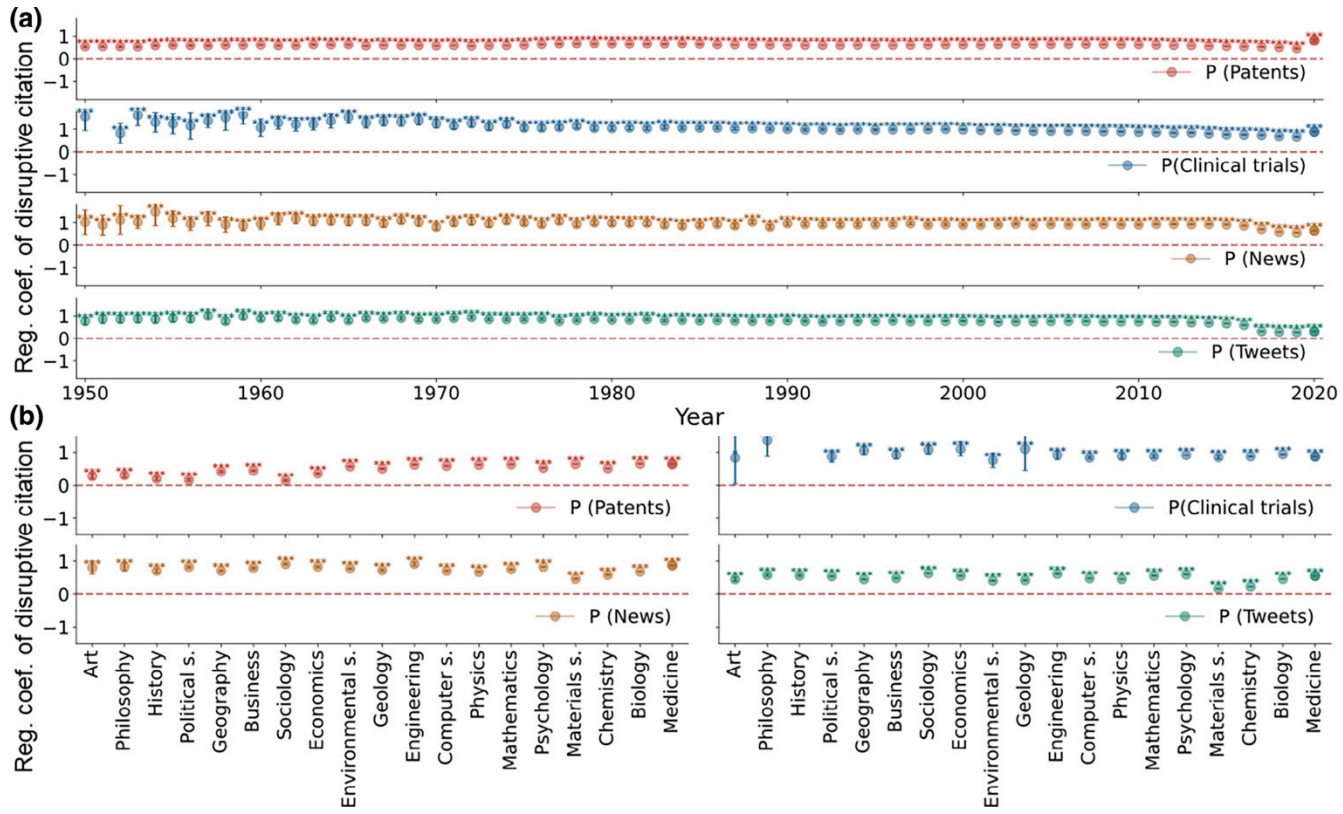


FIGURE 5 Heterogeneity of the relationship between disruptive citation and the probability of papers being linked to technological and societal domains. Each scatter plot depicts the logit regression coefficient of the disruptive citation on the likelihood of papers being linked to patents, clinical trials, news, or tweets, with error bars representing 95% confidence intervals based on robust standard errors. All control variables have been included; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

particularly in STEM fields over the past two decades. Exploring the heterogeneity across temporal and disciplinary boundaries reveals intriguing trends. The identification of a transition point in recent decades, signaling a shift from acceptance to bias against disruptive research, hints at evolving attitudes toward unconventional ideas. Moreover, the varying effects across academic fields underscore discipline-specific subtleties in embracing disruptive research. The significant bias observed against the CD index in STEM fields raises pertinent concerns about such biases and their potential ramifications on scientific advancement and innovation.

The innovative identification and quantification of disruptive impact, as characterized by disruptive citation, provide a fresh perspective for understanding the evolution of scientific paradigms (Funk & Owen-Smith, 2017; Kuhn, 1962; Yang, Hu, et al., 2023). Particularly noteworthy is the consistently positive impact of disruptive citations on the likelihood of papers being linked to technological and societal domains, indicating that highly impactful disruptive papers indeed contribute more to technology and society. These findings are pivotal in shaping discussions on scientific evaluation metrics, advocating for a balanced assessment that duly recognizes disruptive contributions.

The implications of this study extend to stakeholders within academia, funding agencies, and policymakers. By elucidating the underrepresentation of disruptive research in its impact on technology and society, our findings offer valuable insights for guiding resource allocations and strategic decision-making processes (Azoulay et al., 2019; Yin et al., 2021). Policymakers and funding agencies are presented with an opportunity to cultivate environments conducive to fostering high-risk, high-reward research endeavors. By implementing mechanisms that incentivize and support unconventional ideas, they can catalyze transformative innovation and propel societal progress.

Our work offers a nuanced perspective on the observed results, prompting critical reflection from multiple vantage points. Primarily, the discernible bias against the CD index underscores the necessity for a reevaluation of existing evaluation frameworks within academia. Institutions tasked with assessing academic impact and performance metrics must reassess their methodologies to ensure a more inclusive evaluation of disruptive

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

contributions (Hicks et al., 2015). This recalibration is imperative for fostering an academic culture that embraces and rewards innovation across all its forms.

Nevertheless, it is imperative to acknowledge the inherent limitations associated with solely relying on the CD index as a metric for gauging disruptive innovation. While the index furnishes a valuable quantitative framework for identifying disruptive potential, it may not comprehensively capture the nuanced impact of disruptive ideas (Leibel & Bornmann, 2023). Indeed, concerns may arise regarding its capacity to encapsulate the broader societal and technological implications of disruptive research. In addressing this concern, our study introduces a novel approach to quantifying disruptive impact through the scrutiny of disruptive citations. By harnessing this complementary metric, we furnish a more holistic understanding of the tangible contributions of disruptive research to technological and societal advancement. This fresh perspective serves to deepen and broaden our comprehension of disruptive innovation, offering insights into its multifaceted dynamics.

6.2 | Limitations and future avenues

While this study offers comprehensive insights into the relationship between scientific disruption and papers' impact on technology and society, several limitations warrant consideration. First, the reliance on quantitative metrics such as the CD index and disruptive citation might overlook nuanced qualitative aspects inherent in disruptive research. Second, this study predominantly focuses on scientific publications and their links to patents, clinical trials, news sources, and social media. However, other crucial facets of societal impact are not thoroughly explored. Additionally, some papers may not directly link to the technology and society domain, but they indirectly relate to them, such as being cited by papers that are then linked to technology and society, thus we cannot fully identify the effect on the unobserved or indirect links. Lastly, while controlling for various factors, this study might not encompass all potential confounding variables affecting the relationship between disruptive research and its societal impact. Unaccounted factors could influence the observed associations.

Future investigations hold promise in illuminating the dynamic landscape shaped by disruptive research, offering deeper insights into its real-world implications. Expanding the scope of societal impact beyond the realms of patents, clinical trials, news sources, and social media to encompass broader implications—such as policy influence, cultural shifts, or economic transformations—would provide a more holistic assessment of disruptive research's societal implications. Addressing the limitations of indirect links between research and societal domains may involve developing innovative approaches for tracing the pathways of influence. Finally, investigating the mechanisms and avenues through which disruptive innovation impacts technology and society assumes paramount importance in advancing scholarly discourse and guiding informed decision-making.

CODE AVAILABILITY

The code is available at https://github.com/AlexJieYang/Disruption_technology_society. Note that one may need Python-3.10 as well as R-4.3.2 to replicate the codes.

FUNDING INFORMATION

The authors acknowledge the support provided by the Open Fund for Innovative Evaluation (#CXPJ2024004) from Fudan University.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY STATEMENT

The MAG data can be downloaded via https://zenodo.org/record/2628216#.Y-7RR_5Bz-g. Other data used in this study can be obtained by making reasonable requests.

ORCID

Alex J. Yang https://orcid.org/0000-0002-1276-0483
Jia Kong https://orcid.org/0000-0002-4285-9636
Sanhong Deng https://orcid.org/0000-0002-6910-3935

REFERENCES

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

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019). Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *The Review of Economic Studies*, *86*(1), 117–152. https://doi.org/10.1093/restud/rdy034

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*, *361*(6408), 1194–1197. https://doi.org/10.1126/science.aav2484

Bower, J. L., & Christensen, C. M. (1995). Disruptive Technologies: Catching the Wave. *Journal of Product Innovation Management*, 1, 75–76.

Bower, J. L., & Christensen, C. M. (1995a). *Disruptive technologies: Catching the wave*. Harvard Business Review.

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bower, J. L., & Christensen, C. M. (1995b). Disruptive technologies: Catching the wave. *Journal of Product Innovation Management*, *1*, 75–76.

Caplan, N. (1979). The two-communities theory and knowledge utilization. *American Behavioral Scientist*, *22*(3), 459–470. https://doi.org/10.1177/000276427902200308

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), e2021636118. https://doi.org/10.1073/pnas.2021636118

Costas, R., de Rijcke, S., & Marres, N. (2021). "Heterogeneous couplings": Operationalizing network perspectives to study science-society interactions through social media metrics. *Journal of the Association for Information Science and Technology*, *72*(5), 595–610. https://doi.org/10.1002/asi.24427

Fleming, L., Greene, H., Li, G., Marx, M., & Yao, D. (2019). Government-funded research increasingly fuels innovation. *Science*, *364*(6446), 1139–1141. https://doi.org/10.1126/science.aaw2373

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*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185

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

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*(1), 414–427. https://doi.org/10.1162/qss_a_00022

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Jasanoff, S. S. (1987). Contested boundaries in policy-relevant science. *Social Studies of Science*, *17*(2), 195–230. https://doi.org/10.1177/030631287017002001

Kreps, S. E., & Kriner, D. L. (2020). Model uncertainty, political contestation, and public trust in science: Evidence from the COVID-19 pandemic. *Science Advances*, *6*(43), eabd4563. https://doi.org/10.1126/sciadv.abd4563

Kuhn, T. S. (1962). Historical structure of scientific discovery. *Science*, *136*(3518), 760–764.

Lee, C., Kogler, D. F., & Lee, D. (2019). Capturing information on technology convergence, international collaboration, and knowledge flow from patent documents: A case of information and communication technology. *Information Processing & Management*, *56*(4), 1576–1591. https://doi.org/10.1016/j.ipm.2018.09.007

Leibel, C., & Bornmann, L. (2023). What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics*, *129*, 601–639. https://doi.org/10.1007/s11192-023-04873-5

Li, D., Azoulay, P., & Sampat, B. N. (2017). The applied value of public investments in biomedical research. *Science*, *356*(6333), 78–81. https://doi.org/10.1126/science.aal0010

Liang, Z., Mao, J., & Li, G. (2023). Bias against scientific novelty: A prepublication perspective. *Journal of the Association for Information Science and Technology*, *74*(1), 99–114. https://doi.org/10.1002/asi.24725

Lin, Y., Evans, J. A., & Wu, L. (2022). New directions in science emerge from disconnection and discord. *Journal of Informetrics*, *16*(1), 101234. https://doi.org/10.1016/j.joi.2021.101234

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

Lin, Z., Yin, Y., Liu, L., & Wang, D. (2023). SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data*, *10*(1), 315. https://doi.org/10.1038/s41597-023-02198-9

Liu, X., Bu, Y., Li, M., & Li, J. (2024). Monodisciplinary collaboration disrupts science more than multidisciplinary collaboration. *Journal of the Association for Information Science and Technology*, *75*(1), 59–78. https://doi.org/10.1002/asi.24840

Marx, M., & Fuegi, A. (2022). Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *Journal of Economics & Management Strategy*, *31*(2), 369–392. https://doi.org/10.1111/jems.12455

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Packalen, M., & Bhattacharya, J. (2020). NIH funding and the pursuit of edge science. *Proceedings of the National Academy of Sciences*, *117*(22), 12011–12016. https://doi.org/10.1073/pnas.1910160117

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

Petersen, A. M., Arroyave, F., & Pavlidis, I. (2023). Methods for measuring social and conceptual dimensions of convergence science. *Research Evaluation*, *32*, 256–272. https://doi.org/10.1093/reseval/rvad020

Ruan, X., Lyu, D., Gong, K., Cheng, Y., & Li, J. (2021). Rethinking the disruption index as a measure of scientific and technological advances. *Technological Forecasting and Social Change*, *172*, 121071. https://doi.org/10.1016/j.techfore.2021.121071

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

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312), aaf5239. https://doi.org/10.1126/science.aaf5239

Stiglitz, J. E. (1999). Knowledge as a global public good. In *Global public goods: International cooperation in the 21st century*. Oxford University Press. https://doi.org/10.1093/0195130529.003.0015

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

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062. https://doi.org/10.1002/asi.23833

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enough. *Quantitative Science Studies*, *1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, *39*(5), 426–431. https://doi.org/10.2307/3109916

Wu, L. F., Wang, D. S., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

Wuestman, M., Hoekman, J., & Frenken, K. (2020). A typology of scientific breakthroughs. *Quantitative Science Studies*, *1*(3), 1203–1222. https://doi.org/10.1162/qss_a_00079

Xu, F., Wu, L., & Evans, J. (2022). Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences*, *119*(23), e2200927119. https://doi.org/10.1073/pnas.2200927119

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

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

Yang, A. J., Gong, H., Wang, Y., Zhang, C., & Deng, S. (2024). 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. (2023). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management*, *60*(5), 103420. https://doi.org/10.1016/j.ipm.2023.103420

Yang, A. J., Wang, Y., Kong, J., Zhang, Q., Hu, H., Wang, H., & Sanhong, D. (2023). The global disruption index (GDI): An incorporation of citation cascades in the disruptive index. *Proceedings of 19th International Society of Scientometrics and Informetrics Conference*.

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

Yin, Y., Gao, J., Jones, B. F., & Wang, D. (2021). Coevolution of policy and science during the pandemic. *Science*, *371*(6525), 128–130. https://doi.org/10.1126/science.abe3084

Zeng, A., Shen, Z. S., Zhou, J. L., Wu, J. S., Fan, Y., Wang, Y. G., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, *714*, 1–73. https://doi.org/10.1016/j.physrep.2017.10.001

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yang, A. J., Yan, X., Hu, H., Hu, H., Kong, J., & Deng, S. (2024). Are disruptive papers more likely to impact technology and society? *Journal of the Association for Information Science and Technology*, 1–17. https://doi.org/10.1002/asi.24947

23301643, 0, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24947 by Alex Yang - Nanjing University , Wiley Online Library on [04/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License