Research Article

Explaining Disparities in Paper Impact across Similar-Level Journals: An Analysis of Sociology Publications

Nannan Zhao, Linzhuo Li*

Department of Sociology, Zhejiang University, Hangzhou, CHINA.

ABSTRACT

Why do papers of seemingly comparable quality generate widely different levels of scholarly impact? This study addresses this unresolved question by examining internal knowledge characteristics that may explain citation disparities among articles published in three top sociology journals between 1999 and 2022. Drawing on bibliometric techniques, we analyze a dataset of 3,190 articles using Latent Dirichlet allocation (LDA) to extract topic-based representations of each paper. We construct measures of context novelty, content novelty, and knowledge focus, and employ OLS regression models to assess their relationship with five-year citation counts. The analysis was conducted using Python. Contrary to prior findings, novelty measures fail to significantly predict citation outcomes when journal-level prestige is held constant. In contrast, knowledge focus-indicating the thematic concentration of a paper-shows a robust and consistent positive association with citation counts, suggesting that cognitive coherence may enhance scholarly visibility even in structurally comparable settings. This study contributes to ongoing debates about what drives scientific recognition by shifting the focus from external prestige signals to internal cognitive features. It also demonstrates the value of analyzing impact variation within similarly ranked journals-a context often overlooked in large-scale citation studies. The study is limited to the field of sociology, and does not account for factors such as network embeddedness or institutional affiliation. Future research could apply similar designs to other disciplines and incorporate additional contextual variables.

Keywords: Internal characteristics, Content novelty, Context novelty, Knowledge focus, Citation impact, Similar-level journals.

Correspondence:
Linzhuo Li
Department of Sociology, Zhejiang University, Hangzhou, CHINA.
Email: linzhuoli@zju.edu.cn
ORCID: 0000-0001-5732-3642

INTRODUCTION

Scientific impact has long been a focal point in the field of science studies,[1] generating extensive research on its measurement and determinants.[2-4] In the field of science studies, impact frequently operationalized through citations, represents the extent to which a paper influences human intellect or cognition[3] and signifies the dissemination and expansion of knowledge.[5] Given its close relationship with awards, honors, and career progression for scientists,[6,7] journal rankings and reputation for institutions,[7-9] resource allocation for policymakers,[10] and national scientific prestige,[11] the influence of impact transcends individual papers.[12]

Despite the abundance of literature exploring citation-based impact, an essential yet inadequately addressed problem remains: why do papers under similar prestige (e.g., published within similar-level journals) often exhibit significant disparities in their citation counts? Prior research has predominantly examined external institutional factors such as journal reputation, ranking, disciplinary paradigm development etc., are primarily tied to the visibility and social recognition of a paper, highlighting a consensus mechanism that enhances visibility and impact through established reputations.[13-16] However, these external factors alone are insufficient for explaining citation variations among papers under equivalent reputation and ranking. Furthermore, these external indicators typically become accessible only after a paper's publication, limiting their usefulness in predicting paper impact beforehand.

Previous studies indeed have suggested internal indicators, such as paper type, length, reference structure, novelty, immediacy, recency etc., relate to the content characteristics of the paper itself, as influential predictors of scientific impact.[17-21] However, a common limitation across these studies is their overly broad comparative scope, often confounding paper quality and social reputation mechanisms. It remains unclear whether internal characteristics genuinely predict impact or simply reflect structural advantages rooted in journal reputation, disciplinary visibility, or author status. In other words, when external

"consensus-based" mechanisms dominate, the explanatory power of internal features may be overstated.

This paper argues that the variations in impact among papers of similar quality cannot be attributed solely to randomness, in contrast to the social interaction mechanisms proposed in classic sociological studies, such as those examined in music lab experiments, nor to the cumulative advantage of scientific impact driven by prestige-based or consensus-based mechanisms. Instead, it aims to provide an endogenous explanation by analyzing impact variations among papers of similar quality based on their internal characteristics. Using journal articles from the sociology field with comparable ranking levels as a case study, the research extracts intrinsic knowledge features of these papers and specifically examines the effect on citations. This study adopts a cognitive perspective, aiming to explain citation disparities within journals of equal prestige by focusing on how authors make micro-level strategic decisions in assembling and structuring knowledge. Accordingly, we select three internal mechanisms-context novelty, content novelty, and knowledge focus-which capture how authors cognitively structure their work: how disparate knowledge is connected, how thematic content is recombined, and how narrowly the intellectual content is focused. The findings reveal that while novelty does not effectively differentiate citation counts among such papers, the degree of knowledge focus better accounts for the observed impact variations.

LITERATURE REVIEW

Which papers are more likely to be cited? A common assumption is that high-quality papers receive more citations.[22] However, there is no formulaic approach to definitively assess the quality of scientific research.[23] Due to its relative objectivity, citation count has become a widely used metric for evaluating paper quality.[3,24,25] Highly cited papers are often associated with greater creativity and significance,[26-28] and citation advantages tend to accumulate over time, a phenomenon widely known as the Matthew Effect in scientometrics.[5,29,25] This effect is largely driven by reputation-based mechanisms, where heightened visibility enhances the probability of further citations.[15]

Existing studies has predominantly examined external institutional factors, highlighting the role of reputation in shaping citation impact. Academic evaluation systems such as impact factors, the H-index, and CiteScore institutionalize reputation-based assessments. In reputation-based system, journal reputation, in particular, has been found to be a strong predictor of citation counts.[30] Papers published in high-impact journals are more likely to attract citations due to increased visibility,[17-19] and mainstream journals generally exhibit higher citation rates than their less prestigious counterparts.[13] Author reputation also plays a crucial role, as papers by well-established

scholars are more frequently cited.[13,17,18,31,32] Disciplinary differences further influence citation patterns-fields with more established paradigms tend to attract higher citations, as seen in the greater visibility of natural sciences compared to social sciences.[15] Research fields with highly developed paradigms not only receive more resource support but also benefit from faster peer review and shorter publication cycles.[16,33]

Despite the predictive power of reputation-based metrics, these systems primarily serve as post-publication evaluation tools, meaning that citation potential can only be assessed retrospectively, offering little predictive value before publication. Furthermore, citation counts are not perfect indicators of quality but rather reflections of perceived importance, constrained by institutional and social mechanisms.[24] Given these limitations, the challenge remains: how can the impact of a paper be predicted before publication, independent of external reputation signals?

Some scholars argue that citation potential is embedded in a paper's intrinsic attributes rather than external reputation.[17] Factors such as paper length, document type, and reference patterns have been shown to influence citation counts. For instance, longer papers generally attract more citations,[17-19] and review articles tend to be cited more frequently than empirical studies.[18] Structural components, including Figures and Tables, also contribute to a paper's visibility and impact.[20,21] Additionally, citation patterns play a crucial role-papers that incorporate more recent and authoritative references often achieve higher citation counts.[17-19,34,49]

Beyond the formal features above, the role of innovation in citation impact has drawn increasing attention. Creativity is often linked to paper quality, and highly novel papers are more likely to achieve substantial impact.[35] Two dominant paradigms of innovation-combinatorial innovation and disruptive innovation-have been proposed to explain variation in scientific influence. Disruptive innovation is typically assessed based on a paper's long-term effect on existing knowledge structures.[36] Specifically, the disruptive index captures whether subsequent literature continues to cite a focal paper while disregarding its predecessors, making it a fundamentally retrospective metric. As such, it relies on post-publication citation dynamics and is less suitable for studies aiming to predict scientific impact based on pre-publication internal characteristics.

In contrast, combinatorial innovation focuses on the production process itself, arguing that certain knowledge combinations are more innovative than others.[37] Uzzi introduced combinatorial innovation to scientometrics, developing the widely adopted concept of combinatorial novelty by analyzing journal citation patterns.[38] Their findings suggest that impactful scientific contributions often emerge from combinations of highly traditional knowledge with novel, unexpected elements. This combinatorial approach has been found to increase the probability

of a paper becoming a high-impact publication,[39] a conclusion supported by subsequent research.[37]

However, novelty alone may not fully capture the internal cognitive structure of a paper. While novelty emphasizes the distance or atypicality of knowledge combinations, it does not account for the depth or coherence with which ideas are developed and integrated. This aspect of cognitive structure-depth-is closely related to the notion of specialization, which reflects the extent to which research is grounded in a focused and consistent knowledge domain. A growing body of research has examined specialization as a key dimension of scholarly careers, primarily at the disciplinary or author level. These studies explore how scholar's alignment with specific fields or subfields shapes their professional development. Greater specialization is often associated with higher levels of expertise, clearer thematic positioning, and more stable career trajectories.[40-45] Furthermore, focus-as a strategy of specialization-suggests a higher likelihood of success. For example, more focused knowledge is more likely to stand out in the field,[46] promote research productivity,[40] and even lead to groundbreaking or disruptive outcomes.[47] These findings suggest that focus is not merely a stylistic feature of scholarship, but potentially a cognitive mechanism shaping scientific recognition. However, these studies aimed at the level of individual researchers, few have operationalized at paper-level knowledge focus as a measurable construct or tested its relationship with scholarly impact. Yet it still remains unclear how such cognitive strategies manifest at the level of individual publications, especially among papers that are published under similar institutional and disciplinary conditions.

In summary, prior research has identified a wide range of factors influencing scientific impact, which can be broadly categorized into external characteristics and internal characteristics. While external factors are well-established predictors of citation performance, they primarily operate through reputation-based or consensus-driven mechanisms, and as such, can only be observed after a paper is published. Studies of pre-publication internal factors although works, however, they often confounding paper quality and social reputation mechanisms due to their overly broad comparative scope. As a result, existing research has not adequately distinguished how specific internal characteristics contribute independently to variations in citation impact within closely matched journal contexts, highlighting the need to explore those factors that can effectively explain impact differences within micro-contexts.

To address this gap, the study focuses on articles published in three leading journals of comparable reputation in a single field of sociology-American Journal of Sociology, American Sociological Review, and Social Forces. Anchored in a cognitive framework, the study examines three internal characteristics: context novelty, content novelty, and knowledge focus, which capture how authors cognitively structure their work, and investigates whether and how these cognitive-level features can predict differential citation outcomes under conditions of structural equivalence. Our findings indicate that neither form of novelty effectively differentiates citation impact among papers published in journals of comparable ranking. In contrast, knowledge focus proves to be an effective predictor of citation variation, suggesting that even when quality differences are difficult to discern, focus can meaningfully shape scholarly recognition.

METHODOLOGY

As identified in the literature review, the empirical question guiding this study is: Among papers published in journals of similar disciplinary standing, to what extent do internal characteristics-specifically content novelty, context novelty, and knowledge focus-predict differences in citation impact? To address this question, we detail below the data sources and procedures used in this study, including data collection and preprocessing, variable measurement, and the analytical strategy.

Data Collection and Preprocessing

As previously discussed, this study aims to explain variations in impact among papers of comparable quality. However, existing research based on reputation-driven or consensus-based external characteristics-which are only observable through post-publication evaluations-as well as studies relying on pre-publication external indicators-which often conflate paper quality with social reputation mechanisms-are limited in their ability to address this question. To overcome these limitations, we deliberately control for disciplinary context and journal prestige in our data selection. Specifically, we focus on three journals within a single discipline-sociology-namely, American Journal of Sociology, American Sociological Review, and Social Forces-that are of similar ranking or prestige level, as a representative example. This design allows us to examine differences in scholarly impact and the underlying mechanisms under conditions where social structural factors are held constant.

The data used in this study were primarily drawn from the Web of Science (WoS) and the Microsoft Academic Graph (MAG), two widely recognized databases of scholarly publications. We retrieved all research articles information published between 1999 and 2022 [1] by conducting journal-specific searches in WoS. We extracted metadata such as year, titles, abstracts, keywords, and DOIs from WoS. We then manually search and download the content texts of these papers to facilitate text analysis using open-access sources and institutional subscriptions where available based on information of titles and DOIs. In total, we obtained 3,190 papers: 752 from American Journal of Sociology, 931 from American Sociological Review, and 1,507 from Social Forces (see Table 1). In addition, we also collected the historical impact factor data of the three journals from the WoS platform.

1 The starting year of 1999 was selected because these journals began to be indexed more consistently and systematically in WoS from that point onward.

Table 1: Journal-specific data collection details from WoS.

Journal Name	Search Method	Time Span	Number of Articles
American Journal of Sociology	Journal-specific search	1999-2022	752
American Sociological Review	Journal-specific search	1999-2022	931
Social Forces	Journal-specific search	1999-2022	1507

Citation data were obtained by matching each paper's DOI with the Microsoft Academic Graph (MAG), which contains comprehensive citation records across disciplines.

Before analysis, the content texts of each paper were batch-converted to editable .txt files, from which we parsed structural components such as titles, abstracts, main body texts, and footnotes. For the purpose of capturing content-level features, we focused our analysis on the main body of each article, excluding appendices and references. Figure 1 presents an overview of the data collection and preprocessing procedure employed in this study.

Variable Measurement

This study aims to investigate how internal characteristics of scholarly papers contribute to variations in citation impact-especially under conditions where external structural factors such as journal prestige or disciplinary scope are relatively comparable. By focusing on papers published in journals of similar rank and field, we can examine whether internal features can still meaningfully account for differences in scholarly visibility once broader institutional or reputational influences are held constant. Prior studies have identified a range of internal features that tend to correlate positively with citation counts. However, much of this evidence comes from heterogeneous samples across fields and journal tiers, where internal features are often confounded with paper quality or reputation-based mechanisms. As a result, it remains unclear whether such internal attributes retain their predictive value in more structurally uniform contexts.

To address this question, we test the extent to which three specific internal characteristics can distinguish impact among papers published in journals of comparable standing. While this selection inevitably excludes other potentially relevant features, we believe that focusing on context novelty, content novelty, and knowledge focus is beneficial, as these three factors capture deeper cognitive dimensions of scholarly work. That is, they speak directly to how authors organize, integrate, and position knowledge within the intellectual space: how disparate ideas are connected, how thematic content is recombined, and how narrowly the argument is focused. Specifically, context novelty measures the uniqueness of a paper's knowledge base, based on how typical its cited journal combinations are; content novelty captures the rarity of thematic combinations, derived from topic modeling; and focus assesses the degree of concentration in a paper's topic distribution. The measurement details are outlined below.

Context novelty

Context novelty captures the degree to which a paper draws on atypical combinations of prior sources-specifically, the rarity of journal pairings in its reference list. This measure reflects how distinctive or unconventional a paper's knowledge base is in relation to prevailing citation patterns. Specifically, we use commonness as a proxy variable for context novelty. Following previous study,[37] we define the commonness of each journal pair (journals i and j) in year t as follows:

$$Commoness_{ijt} = \frac{N_{\text{observed journal pairs } ijt}}{N_{\text{expected journal pairs } ijt}} = \frac{N_{ijt}}{\frac{N_{it}}{N_t} \cdot \frac{N_{jt}}{N_t} \cdot N_t} = \frac{N_{ijt} \cdot N_t}{N_{it} \cdot N_{jt}}$$

We identify all journal pairs cited by a given paper and calculate their commonness based on how frequently each pair co-occurs in other papers published in the same year. We define the reference universe as all papers in the sample published in year t, and calculate the expected co-occurrence of each journal pair (i, j) based on their marginal frequencies. The observed-over-expected ratio indicates how typical or atypical the journal pairing is. For each paper, we assign commonness scores to all cited journal pairs, then take the 10th percentile of these values as a conservative estimate of contextual distinctiveness. This value is log-transformed and sign-reversed so that higher scores reflect greater context novelty. Intuitively, a paper citing highly uncommon journal combinations is considered to possess a more original knowledge foundation.

Content novelty

Content novelty captures the originality of a paper's thematic structure by assessing the rarity of its core topic combinations relative to recent literature, which distinguishes content-level novelty from citation-based indicators by focusing on the internal composition of ideas. The underlying premise is that papers recombining commonly separated topics may introduce novel conceptual linkage or problem framing, thereby increasing their visibility and impact. Content novelty is defined as the rarity or uniqueness of the combination of two representative topics:

$$Novelty = 1 - r$$

and,

$$r = \frac{N_{\text{observed topic } ij}}{N_{\text{expected topic } ij}} = \frac{N_{ij}}{\frac{N_i}{N} \cdot \frac{N_j}{N} \cdot N} = \frac{N_{ij} \cdot N}{N_i \cdot N_j}$$

Specifically, using paper text from the three target journals, we trained an LDA topic model and identified the two most prominent topics for each paper based on topic loadings, we then

computed the co-occurrence frequency of selected topic pairs within a three-year window preceding each paper's publication. The rarity of a topic combination was calculated by comparing its observed frequency to the expected frequency under an assumption of independence. Lower observed-over-expected ratios indicate more novel thematic pairings.

Focus

Knowledge focus captures the degree to which a paper concentrates its intellectual content within a narrow thematic scope, as opposed to distributing its content broadly across multiple thematic areas. The underlying idea is that a more focused paper may offer greater conceptual clarity, stronger domain specificity, and thus more easily identifiable scholarly contributions.[40,42]

We operationalize focus using the Herfindahl-Hirschman Index (HHI), a widely used measure of concentration originally developed in economics, as:

$$\text{Focus} = \sum_{i}^{k} \theta_i^2$$

In our context, it reflects how unevenly a paper distributes its content across different topics. For each paper, we take the posterior topic distribution produced by the LDA model and compute the sum of the squared topic loadings. This yields a score ranging from 0 to 1, where values closer to 1 indicate a higher concentration of content in fewer topics (i.e., greater knowledge focus), and values closer to 0 indicate a more evenly distributed topic profile (i.e., less focus).

Analytical Strategy

To examine the relationship between internal characteristics and scholarly impact, we employ an Ordinary Least Squares (OLS) regression model, with the number of citations a paper receives within five years of publication as the dependent variable. This five-year citation window is commonly used in bibliometric research[2] to capture a paper's medium-term influence, and we also tried other window (like 10 years). The three focal independent variables-content novelty, context novelty, and knowledge focus-represent distinct dimensions of cognitive structuring. To account for temporal variation in citation practices and exposure time, we include publication year fixed effects as control variables. We also include team size as a control variable to partially account for author-level effects such as visibility or collaboration-based advantages, which may be
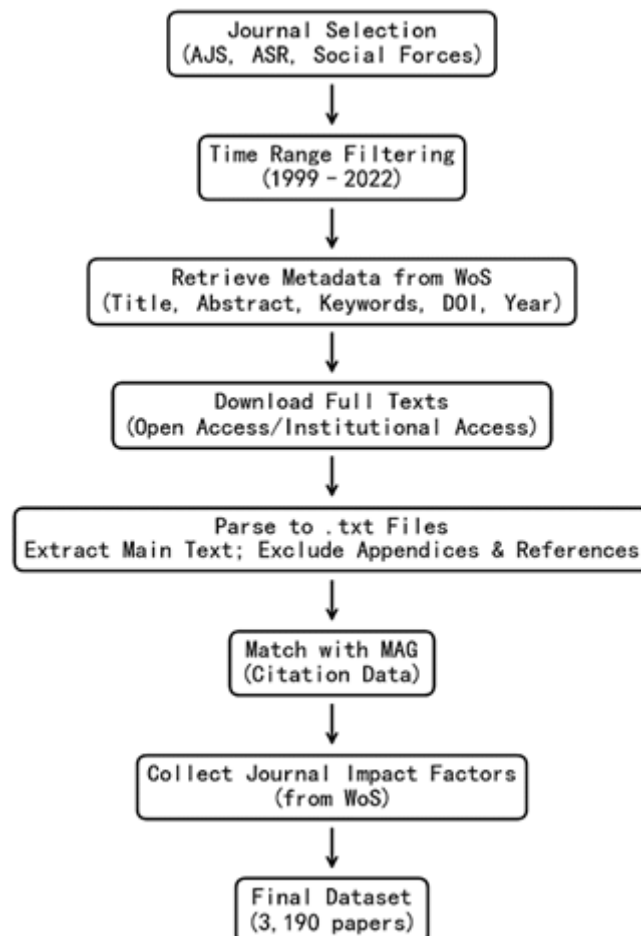
Figure 1: Data collection and preprocessing workflow.

correlated with reputation or institutional standing. In addition, we also conducted some robustness checks.

RESULTS

Exploratory Patterns in Journal Impact Factors

As detailed in the Data Collection and Preprocessing section, journal impact factor data for the three selected journals were retrieved from the Web of Science database. Figure 2 illustrates the changes in impact factors for the three journals from 1999 to 2022. While all three journals exhibit an overall upward trajectory, notable differences emerge over time. The American Sociological Review (ASR) and Social Forces show more substantial growth-particularly after 2015-whereas the American Journal of Sociology (AJS) displays a slower and more gradual increase. Before 2013, AJS and ASR had comparable impact factors, both consistently higher than Social Forces. However, in subsequent years, ASR pulled ahead, while Social Forces gradually closed the gap with AJS, even surpassing it in recent years.

Although all three journals are considered top-tier outlets in sociology, this divergence raises questions about whether internal publication characteristics-beyond journal reputation-may account for differential citation dynamics. In the following sections, we explore this possibility by examining the internal cognitive features of individual papers across these journals.

Table 2: Topics and representative vocabulary.

Topics	Representative Vocabulary
Law/Crime	Criminal, crime, incarceration, law, justice, legal, discrimination, court, prison, record, punishment, arrest, police, sanctions, drug, treatment, judges.
Education	Students, schools, parents, high school, achievement, academic, teachers, grade, schooling, parental, colleges, score, graduates, enrollment, peers.
Social Movement	Protest, participants, civic, movement, legitimacy, media, event, perceptions, actor, cooperation, resource, norm, emotions, activists, leaders, authority.
Community/ Segregation	Neighborhood, segregation, poverty, residents, city, census, crime, blacks, spatial, moving, urban, metropolitan, residential, disadvantage, concentration.
Migration	Immigrants, migration, migrants, born, generation, south, foreign, assimilation, native, origin, language, identity, Latino, Mexican, blacks.
Religion	Religious, religion, church, religiosity, attendance, protestant, catholic, Muslim, beliefs, christian, moral, caste, god, secular, pornography, Islamic.
Sex/Gender	Sexual, sex, masculinity, domains, gay, identity, love, couples, HIV, gendered, lesbian, sexual harassment, heterosexual, ideology, abortion.
Election	Tax, clients, money, financial, credit, elections, voters, voting, asylum, welfare, contributions, party, reform, bankruptcy, capitalists, crisis, prices.
Network/Capital	Network, ties, friends, social capital, peer, connected, contact, clustering, interpersonal, cohesion, formation, density, attachment, small word, weak.
Party/Politics	Party, rights, civil, elite, military, politics, democratic, conflict, reform, revolution, radical, congress, regime, nation, city, grievances, opposition.
Occupation/ Employment	Wage, earnings, career, worker, unemployment, skill, jobs, post, labor market, occupations, welfare, income inequality, sector, transition, loss.
Racial/Discrimination	Discrimination, African, whites, color, skin, Americans, African Americans, residents, multiracial, racism, conflict, biracial, comments, disparities.
Nation/Globalization	Global, international, environmental, globalization, foreign, nations, trade, domestic, investment, democracy, cross national, regional, diffusion.
Life Course/Mental Health	Mental, birth, cohort, parental, adults, life course, childhood, genetic, exposure, ages, depression, stress, well-being, adolescent, young adult.
Marriage/Family	Mothers, child, marriage, fathers, marital, couples, divorce, care, birth, childcare, fertility, housework, maternal, happiness, partners, cohabitation.
Economics/ Organization	Industry, firms, business, corporate, union, managers, employees, diversity, financial, markets, governance, workplace, executive, ownership, products.
Violence/Terrorism	Police, violence, gun, policing, law, crime, hate, attacks, aids, enforcement, chiefs, event, terrorism, threat, hybrid, lynching, legal, transnational.

Table 3: Pearson Correlation Matrix of Independent Variables.

	Context novelty	Content novelty	Focus
Context novelty	1	0.0256	0.0121
Content novelty	0.0256	1	-0.0033
Focus	0.0121	-0.0033	1

Table 4: Variance Inflation Factors (VIF) of Independent Variables.

Variable	VIF
Constant (intercept)	17.037903
Context novelty	1.000806
Content novelty	1.000671
Focus	1.000158

Internal characteristics

Topic Structure and Distribution Patterns

Before analyzing the topical composition of the journals, we evaluated the coherence scores of LDA models with varying numbers of topics (ranging from 5 to 40). As shown in Figure 3, coherence peaks at 25 topics (0.629), but remains relatively stable. While 25 topics offered the highest quantitative score, we also considered semantic interpretability through manual inspection of topic-word distributions. A 20-topic solution was selected as it provided a balance between thematic clarity and model coherence, facilitating both analytical tractability and conceptual relevance.

Table 2 presents the 20 extracted topics along with their most representative vocabulary terms (only displaying the topics with clear significance). These topics reflect major thematic domains within sociology, including law and crime, education, social movements, religion, gender, and more.

Figure 4 illustrates the temporal distribution of topics across the three journals. ASR and Social Forces exhibit relatively balanced topic coverage, with no sustained dominance of particular themes. In contrast, AJS demonstrates a stronger and more persistent emphasis on certain topics-especially social movements and, during 2011-2017, violence and terrorism. This thematic concentration may partially contribute to the journal's slower growth in impact factor over the years. However, this observation remains tentative, as the relationship between topical focus and journal-level citation outcomes is complex and likely influenced by multiple factors. It is also important to distinguish journal-level topic patterns from paper-level knowledge focus, which is the main object of analysis in subsequent sections.

Importantly, this journal-level pattern does not contradict our subsequent findings at the individual-paper level, where we observe that greater knowledge focus-the degree to which a paper concentrates its thematic content-can enhance citation impact among papers of comparable journal ranking. In other words, while excessive topical repetition may limit a journal's reach, cognitive focus at the article level may help a paper stand out in a crowded intellectual space.

Content and context novelty

Figure 5 shows the variations in content novelty and context novelty for the three journals. Regarding content novelty, there are significant differences between the three journals. The content novelty of AJS is lower than the other two journals for most of the time, especially lower than ASR. The content novelty of AJS and Social Forces fluctuates more widely, while ASR's content novelty changes relatively steadily. In addition, the content novelty of AJS and Social Forces shows considerable fluctuation, while ASR remains relatively stable with a slight increase throughout the observed period. For context novelty, AJS is higher than the other two journals in most cases, particularly higher than Social Forces. At the same time, AJS's context novelty fluctuates significantly more than the other two journals. AJS and ASR show significant peaks in context novelty in some years, while Social Forces remains relatively stable.

Furthermore, consistent with previous research,[48] the correlation between the two types of novelty we measured is weak (coefficients<0.03), supporting the idea that they capture distinct dimensions of innovation. These indicators will be analyzed further in the regression models to assess their predictive value for citation impact.

Focus

Figure 6 displays the temporal patterns of knowledge focus across the three journals. ASR maintains the most stable and consistently high level of focus, suggesting that its published papers are more thematically concentrated. In contrast, AJS exhibits a gradual decline in focus-especially between 2007 and 2013-and remains lower overall than the other two journals. Social Forces shows the greatest volatility, with substantial year-to-year fluctuations across the observation period. Whether such journal-level patterns translate into citation differences at the paper level is addressed in the regression analysis that follows.

Model results

Collinearity Diagnostics and Descriptive statistics

To examine how internal characteristics of papers shape their scholarly impact, we employ an Ordinary Least Squares (OLS) regression model. Following previous work,[18,36,49] we model five-year citation counts as a function of the three main internal characteristics: context novelty, content novelty, and focus.

Before conducting the regression analysis, we performed multicollinearity diagnostics for the main explanatory variables. we conducted collinearity diagnostics using two methods. First, a Pearson correlation matrix (see Table 3) shows that all pairwise correlations among the three main predictors are below 0.03, indicating extremely weak linear relationships. Second, we computed Variance Inflation Factors (VIF), all of which are approximately 1.0-well below conventional thresholds for multicollinearity concerns (see Table 4). These results confirm that the three predictors capture conceptually distinct dimensions of internal structure.

Table 5 presents the descriptive statistics for both the independent and dependent variables. To construct the analytical sample, we first matched the records from the three sociology journals with citation and reference data from the Microsoft Academic Graph (MAG), yielding an initial sample of 1,117 papers. Since the dependent variable is citation count within five years of publication, we excluded papers published after 2015 to ensure sufficient time for citation accumulation. In addition, content novelty is calculated based on the rarity of a paper's top two topic combinations compared to those in the previous three years. Given that our dataset starts in 1999, we could only compute content novelty for papers published in 2002 or later. As a result, papers published before 2002 were excluded from the final model. After applying these restrictions, the final sample size used for regression analysis is 967 observations.

Regression model results

Table 6 presents the OLS model results for the individual characteristics of the papers and their citations. In this model, as mentioned earlier, we use the citation count within five years after publication as the dependent variable (we also tried 10-year window), with content novelty, context novelty, focus as the main explanatory variables, and control for year. We also include team size as a control variable to partially account for author-level effects, which may be correlated with reputation or institutional standing. The detailed results of the main model are shown in Table 5.

From Table 6, we observe that neither content novelty nor context novelty has a statistically significant effect on five-year and ten-year citation counts after publication. This contrasts with findings from previous studies that have associated higher novelty with greater impact and visibility.[38,48,50] One possible reason for this divergence is that, when controlling for structural comparability across similarly ranked journals, the marginal advantage of novelty may be diminished.

In contrast, knowledge focus shows a strong and statistically significant positive association with citation impact ($p<0.01$). Papers that exhibit higher thematic concentration tend to receive more citations, suggesting that cognitive coherence may enhance visibility even in settings where external prestige signals are held constant. While this result underscores the potential value of
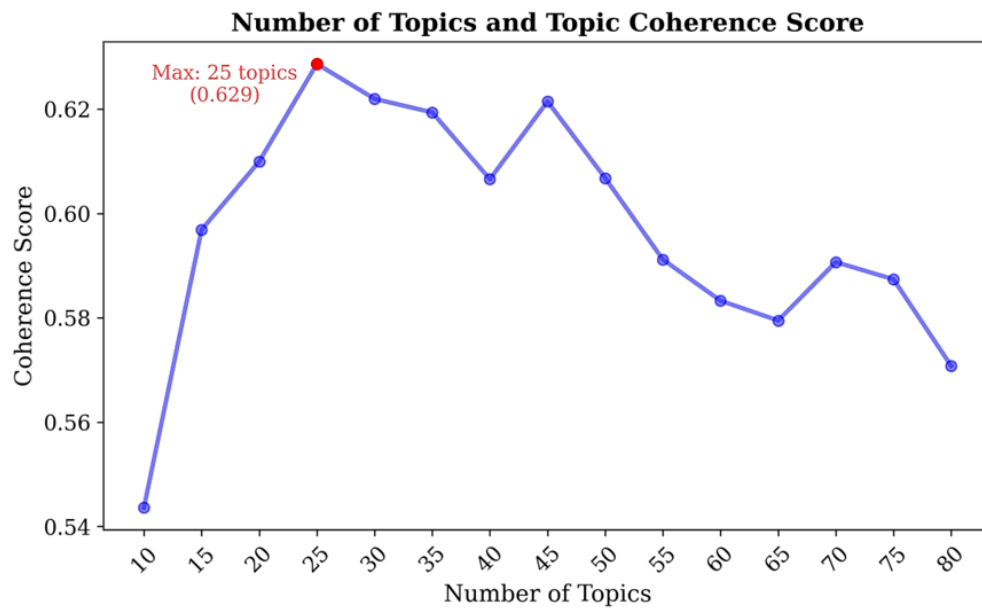
Table 5: Descriptive statistics of the variables.

Variable	Obs	Mean	Std	Min	Max
CiteIn5Y	967	29.30	38.02	0	464
Content novelty	967	-0.92	2.16	-24.09	1
Context novelty	967	-3.99	0.48	-4.53	-1.66
Focus	967	0.35	0.14	0.12	0.99
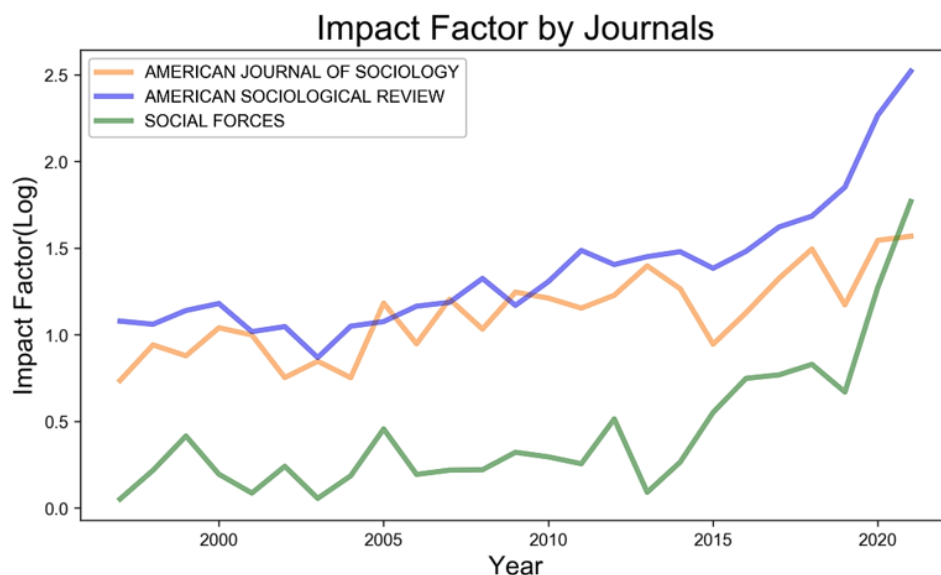
Figure 2: Changes in impact factors.

Figure 3: Number of topics and topic coherence score.

Table 6: OLS model results of novelty, focus on 5-year and 10-year citations.

Y=CiteIn5Y	Model (20 topics)
Content novelty	0.0066（0.014）
Context novelty	0.0715（0.064）
Focus	0.7202**(0.219)
Year	controlled
Team size	controlled
R^2	0.295
N	967
Y = CiteIn10Y	**Model (20 topics)**
Content novelty	0.0099（0.015）
Context novelty	0.0739（0.065）
Focus	0.7419**(0.224)
Year	controlled
Team size	controlled
R^2	0.458
N	967

*$p<0.05$, **$p<0.01$, ***$p<0.001$ (two-tailed t test).

focus as an internal quality indicator, further research is needed to explore how it interacts with other mechanisms such as author reputation or network position.

Robustness Checks and Sensitivity Analysis

To assess the robustness of our findings, we re-estimated the models using alternative topic model specifications (15 and 30 topics). As shown in Table 7, knowledge focus continues to exhibit a statistically significant positive relationship with citation counts, regardless of topic number. In contrast, the effects of content and

context novelty remain largely non-significant, even with only a marginal negative coefficient for content novelty. As a further sensitivity analysis, we also extended the citation window from five to ten years. This longer-term measure helps assess whether the observed patterns persist beyond initial post-publication dynamics. The results remain largely consistent: knowledge focus retains a significant positive effect, while novelty measures again show limited predictive value. Overall, these analyses reinforce the robustness of our main findings. The positive association between thematic concentration and citation impact holds across topic models and time windows, suggesting that focus operates as a stable internal characteristic may associated with greater scholarly visibility in similarly ranked journals.

In summary, our results suggest that the distinctiveness of a paper's topical focus is positively associated with citation impact-i.e., papers with more concentrated knowledge tend to receive more citations. In contrast, neither context novelty nor content novelty shows a statistically significant effect in any of the model specifications, despite their theoretical prominence in prior literature. The summarized results of the models are provided in Table 8. These findings do not negate the broader value of novelty-based approaches, but indicate that within the controlled context of journals of similar ranking, knowledge focus emerges as a more robust internal predictor of citation differences. Further research could explore whether the effects of novelty operate differently across fields or in less structurally comparable publication settings.

LIMITATIONS

While our analysis concentrates primarily on the internal cognitive features of academic papers-such as their conceptual focus and thematic novelty-we fully acknowledge that citation

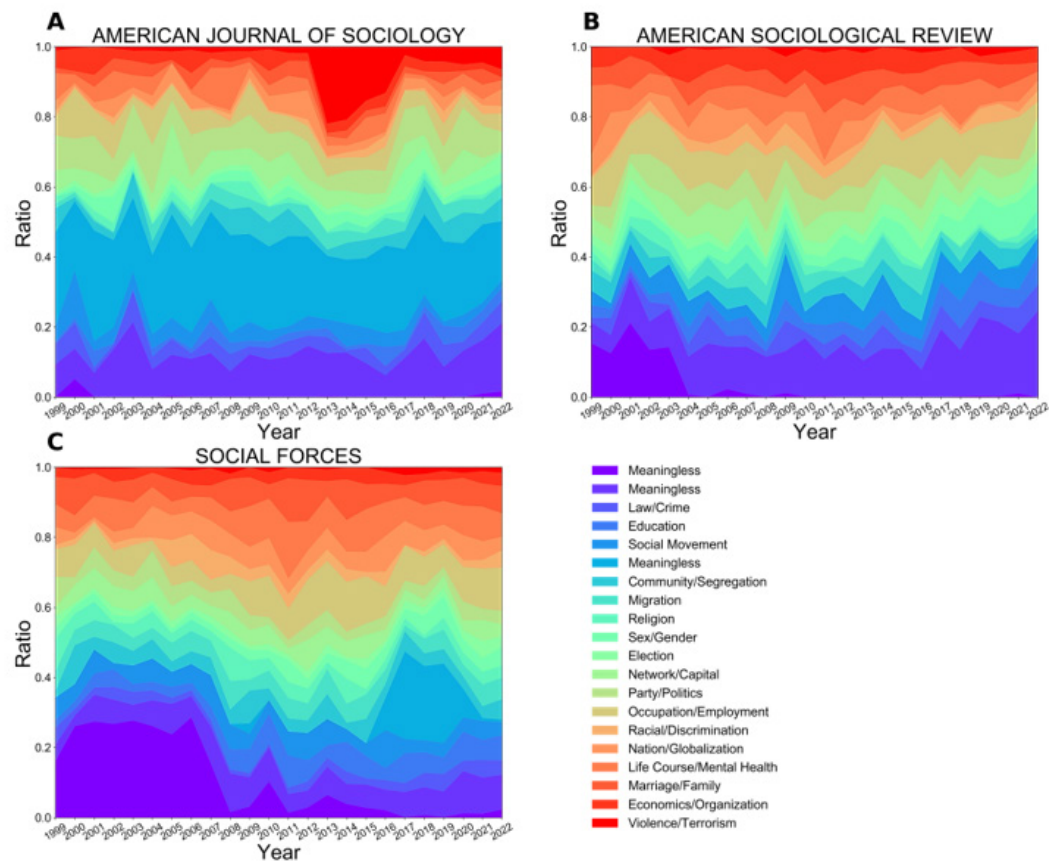## Topic Distribution Over Years by Journals

Figure 4: Topic distribution over years.

Table 7: OLS model results under different numbers of topics.

Y=CiteIn5Y	Model 1 (15 topics)	Model 2 (30 topics)
Content novelty	-0.0003(0.021)	-0.0193(0.007)
Context novelty	0.0731(0.064)	0.0482(0.065)
Focus	0.5274**(0.202)	0.5260*(0.225)
Year	controlled	controlled
Team size	controlled	controlled
R²	0.292	0.298
N	967	967
Y=CiteIn10Y	**Model 1 (15 topics)**	**Model 2 (30 topics)**
Content novelty	0.0031(0.0022)	-0.0181(0.007)
Context novelty	0.0764(0.065)	0.0500(0.066)
Focus	0.5250*(0.206)	0.5097*(0.230)
Year	controlled	controlled
Team size	controlled	controlled
R²	0.455	0.458
N	967	967

*$p<0.05$, **$p<0.01$, ***$p<0.001$ (two-tailed t test).

impact is also significantly influenced by a range of external social and institutional factors. These may include, but are not limited to, prevailing disciplinary citation norms, the professional reputation of the authors, the prestige and visibility of their affiliated institutions, and patterns of self-citation. Such factors often operate in subtle and overlapping ways, shaping how scholarly work is received and recognized within academic communities.

To partially control for structural variation across different publication venues, we deliberately limited our sample to articles published in three highly prestigious sociology journals. This sampling strategy serves to mitigate some sources of heterogeneity, especially those related to journal prestige or disciplinary subfields. However, this focus also introduces limitations with respect to the broader applicability of our findings. Specifically, it constrains the generalizability of our conclusions beyond the specific field of sociology, and may not reflect dynamics present in other disciplines or less prominent journals.

Additionally, we rely on topic modeling using Latent Dirichlet allocation (LDA), a widely used but inherently limited computational technique. LDA's outcomes are sensitive to the selection of parameters (such as the number of topics), and the

interpretive process involved in labeling topics introduces a degree of subjective judgment. As a result, the model may not fully capture the nuanced and layered meanings embedded in complex scholarly texts.

DISCUSSION

Prior research on scientific impact has emphasized both external institutional mechanisms-such as journal prestige, author reputation-and internal characteristics of scholarly work, including length, citation density, and novelty. While these factors have offered valuable insights, a persistent question remains: among papers published in similarly prestigious venues, what explains the variation in scholarly impact? In such cases, where structural advantages are relatively controlled, internal cognitive characteristics may play a more consequential role. This study contributes to that line of inquiry by examining how papers' novelty and knowledge focus relate to citation outcomes within three similar-level sociology journals. More broadly, our study underscores the importance of looking beyond traditional prestige-based or network-driven explanations of scientific

recognition. Especially in contexts where external status signals are already saturated, internal cognitive features-such as how ideas are assembled, framed, and integrated-may matter more than previously assumed. This perspective also invites renewed attention to micro-level variation: small differences in how papers articulate knowledge may produce outsized effects in how they are received and cited, even among works of ostensibly similar quality.

Our analysis shows that, under conditions of structural comparability, knowledge focus-as measured by thematic concentration-is a stable and consistent predictor of citation impact, whereas both content and context novelty lose their statistical significance. This finding suggests that, even when author reputation and journal visibility are held relatively constant, how narrowly and coherently a paper organizes its intellectual contribution may enhance its chances of being cited. Focus may improve a paper's conceptual legibility and recognizability within its field, facilitating easier uptake by other researchers. In contrast, the benefits of novelty-particularly when
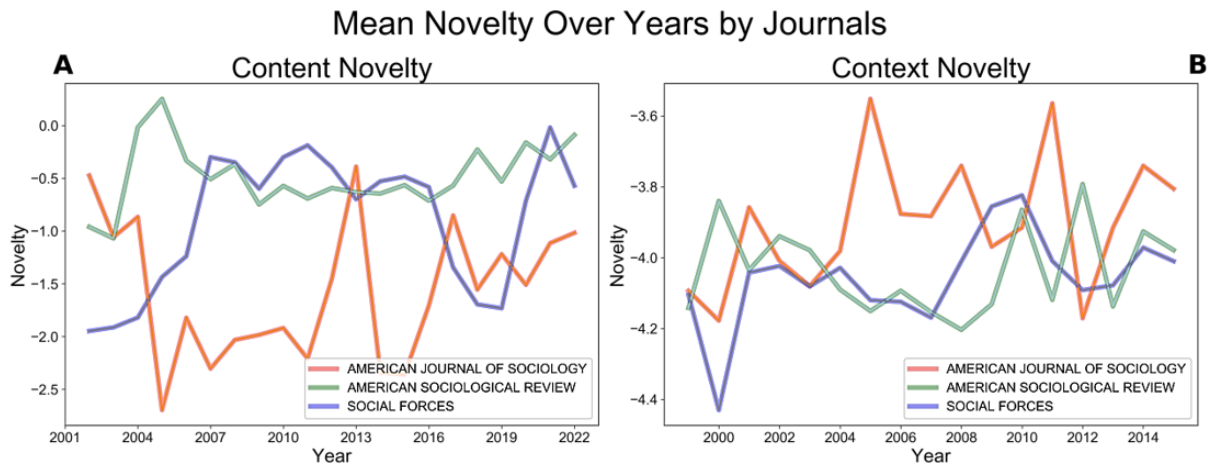
Figure 5: Changes of novelty.

Table 8: Summarized OLS model results.

Model Specification	Citation Window	Content Novelty	Context Novelty	Focus	R²
20 topics (main model)	5 years	0.0066 (n.s.)	0.0715 (n.s.)	0.7202 (**)	0.295
20 topics (main model)	10 years	0.0099 (n.s.)	0.0739 (n.s.)	0.7419 (**)	0.458
15 topics (robustness check)	5 years	-0.0003 (n.s.)	0.0731 (n.s.)	0.5274 (**)	0.292
15 topics (robustness check)	10 years	0.0031 (n.s.)	0.0764 (n.s.)	0.5250 (*)	0.455
30 topics (robustness check)	5 years	-0.0193 (n.s.)	0.0482 (n.s.)	0.5260 (*)	0.298
30 topics (robustness check)	10 years	-0.0181 (n.s.)	0.0500 (n.s.)	0.5097 (*)	0.458

Notes: *p<0.05, **p<0.01, ***p<0.001; n.s.=not statistically significant. All models control for year and team size.

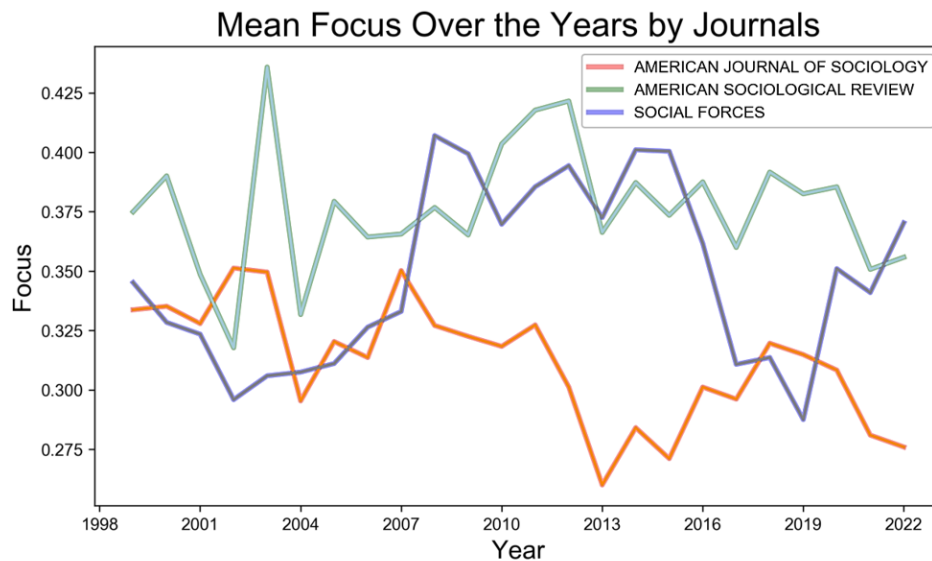## Mean Focus Over the Years by Journals

Figure 6: Changes of knowledge focus.

not accompanied by sufficient cognitive anchoring-may be more limited or delayed.

These findings partially align with, but also diverge from, prior research. For example, previous studies have emphasized the positive association between novelty-particularly atypical combinations-and citation impact.[38,39] However, their analyses typically draw from broader, cross-disciplinary datasets with high variation in journal prestige and author visibility. In contrast, our study holds these structural factors relatively constant, suggesting that the predictive value of novelty may be contingent on variation in social prestige or disciplinary breadth. Previous studies also have shown that the relationship between novelty and impact may follow an inverted-U pattern: moderately novel papers are more likely to be cited, while extremely novel ones may be too disconnected to be immediately recognized.[51-53] Furthermore, highly novel work often carries reputational and interpretive risks-it may be overlooked, misunderstood, or underappreciated, especially in its early stages. While our study does not directly test these mediating mechanisms, our findings are consistent with the idea that in structurally saturated environments, the payoff of novelty may diminish or become more unpredictable. That said, our results do not challenge the theoretical importance of novelty, but instead point to its conditional effects.

On the other hand, our finding that knowledge focus predicts impact resonates with emerging work on thematic coherence and research specialization. For instance, some studies demonstrate that scholars who pursue more focused intellectual trajectories tend to receive greater recognition.[41,42] While most of this literature has focused on the author level, our findings extend this insight to the paper level, showing that coherence within a single publication may matter as much as career-long focus.

Our findings also contribute to understanding the relative changes in journal impact factors. We observed that among the three journals examined, the American Journal of Sociology (AJS) has experienced a slower increase in its impact factor over the past two decades. In parallel, our topic modeling shows that AJS has consistently concentrated on a narrower set of themes-particularly a sustained emphasis on topics related to social movements-compared to ASR and Social Forces. While this overlap in trends is noteworthy, we do not claim a causal relationship between thematic concentration and journal-level impact. These observations may raise a potential hypothesis: that reduced topical diversity may limit a journal's reach across subfields or audiences. However, this proposition remains speculative and would require dedicated empirical testing in future research.

CONCLUSION

This study set out to investigate the extent to which internal cognitive characteristics of scientific papers-namely, context novelty, content novelty, and knowledge focus-affect their subsequent citation impact when published in sociology journals of comparable ranking. While the role of novelty has long been emphasized in the literature as a central driver of scientific influence and recognition, our empirical findings challenge this assumption by showing that its predictive strength weakens in settings where structural factors, such as journal prestige and editorial reputation, are held relatively constant. In other words, novelty alone does not necessarily translate into higher impact when external evaluation cues are uniform across publication venues.

In contrast, knowledge focus-operationalized as the degree of thematic concentration within a paper-consistently and significantly predicts citation performance. This suggests that

when the structural playing field is level, the internal organization of ideas, particularly the coherence and specificity with which a research contribution is articulated, becomes a more decisive factor in gaining scholarly attention. These findings align with and extend a growing body of scholarship that underscores the importance of cognitive structure in shaping how scientific work is received and valued by the academic community. They also contribute to an emerging understanding that scholarly visibility may depend as much on how ideas are framed and presented as on the novelty of the ideas themselves.

Several limitations of this study warrant careful consideration. First, while we aimed to control for journal-level prestige, we could not account for other potentially influential factors such as institutional affiliation, network embeddedness, or prior visibility. Second, although we focused on three high-ranking journals in sociology, the findings may not generalize to fields with different citation practices or epistemic cultures, and thus restrict the broader applicability of our conclusions. Third, our analysis emphasized novelty and focus as internal features, but other cognitive dimensions-such as clarity, methodological rigor, or rhetorical structure-may also shape impact. Future research should expand the scope of internal characteristics and explore how they interact with social mechanisms to influence scholarly visibility. However, ultimately, this study advances our understanding of how impact disparities emerge among papers of comparable quality, and underscores the importance of internal coherence-beyond novelty alone-as a mechanism of scholarly visibility.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

FUNDING

REFERENCES

1. Bu Y, Waltman L, Huang Y. A multidimensional framework for characterizing the citation impact of scientific publications. Quant Sci Stud. 2021; 2(1): 155-83. doi: 10.1162/qss_a_00109.

2. Wang J. Citation time window choice for research impact evaluation. Scientometrics. 2013; 94(3): 851-72. doi: 10.1007/s11192-012-0775-9.

3. Bornmann L, Daniel HD. What do citation counts measure? A review of studies on citing behavior. J Doc. 2008; 64(1): 45-80. doi: 10.1108/00220410810844150.

4. Tahamtan I, Safipour Afshar A, Ahamdzadeh K. Factors affecting number of citations: A comprehensive review of the literature. Scientometrics. 2016; 107(3): 1195-225. doi: 10.1007/s11192-016-1889-2.

5. Merton RK. The Matthew effect in science: the reward and communication systems of science are considered. Science. 1968; 159(3810): 56-63. doi: 10.1126/science.159.3810.56, PMID 5634379.

6. Inhaber H, Przednowek K. Quality of research and Nobel Prizes. Soc Stud Sci. 1976; 6(1): 33-50. doi: 10.1177/030631277600600102.

7. Simons K. The misused impact factor. Science. 2008; 322(5899): 165. doi: 10.1126/science.1165316, PMID 18845714.

8. Anderson RC, Narin F, McAllister P. Publication ratings versus peer ratings of universities. J Am Soc Inf Sci. 1978; 29(2): 91-103. doi: 10.1002/asi.4630290208.

9. Hagstrom WO. Inputs, outputs, and the prestige of university science departments. Sociol Educ. 1971; 44(4): 375-97. doi: 10.2307/2112029.

10. Carlsson H. Allocation of research funds using bibliometric indicators—asset and challenge to Swedish higher education sector. Info. 2009; 64(4): 82-8.

11. King DA. The scientific impact of nations. Nature. 2004; 430(6997): 311-6. doi: 10.1038/430311a, PMID 15254529.

12. Waltman L. A review of the literature on citation impact indicators. J Inf. 2016; 10(2): 365-91. doi: 10.1016/j.joi.2016.02.007.

13. Bornmann L, Schier H, Marx W, Daniel HD. What factors determine citation counts of publications in chemistry besides their quality? J Inf. 2012; 6(1): 11-8. doi: 10.1016/j.joi.2011.08.004.

14. Zuckerman H, Merton RK. Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. Minerva. 1971; 9(1): 66-100. doi: 10.1007/BF01553188.

15. Lodahl JB, Gordon G. Differences between physical and social sciences in university graduate departments. Res Higher Educ. 1973; 1(3): 191-213. doi: 10.1007/BF00991529.

16. Hargens LL. Scholarly consensus and journal rejection rates. Am Sociol Rev. 1988; 53(1): 139-51. doi: 10.2307/2095739.

17. Haslam N, Ban L, Kaufmann L, Loughnan S, Peters K, Whelan J, et al. What makes an article influential? Predicting impact in social and personality psychology. Scientometrics. 2008; 76(1): 169-85. doi: 10.1007/s11192-007-1892-8.

18. Peters HP, van Raan AF. On determinants of citation scores: A case study in chemical engineering. Journal of the American Society for Information Science. 1994; 45(1): 39-49. doi: 10.1002/(SICI)1097-4571(199401)45:1<39::AID-ASI5>3.0.CO;2-Q

19. Mammola S, Fontaneto D, Martínez A, Chichorro F. Impact of the reference list features on the number of citations. Scientometrics. 2021; 126(1): 785-99. doi: 10.1007/s11192-020-03759-0.

20. Cleveland WS. Graphs in scientific publications. Am Stat. 1984; 38(4): 261-9. doi: 10.1080/00031305.1984.10483223.

21. Simonton DK. Scientific status of disciplines, individuals, and ideas: empirical analyses of the potential impact of theory. Rev Gen Psychol. 2006; 10(2): 98-112. doi: 10.1037/1089-2680.10.2.98.

22. Teplitskiy M, Duede E, Menietti M, Lakhani KR 2020. Status drives how we cite: Evidence from thousands of authors. arXiv preprint arXiv:2002.10033.

23. Figueredo E. The numerical equivalence between the impact factor of journals and the quality of the articles. J Am Soc Inf Sci Technol. 2006; 57(11): 1561. doi: 10.1002/asi.20418.

24. Smith LC. Citation analysis. Libr Trends. 1981; 30: 83-106.

25. Garfield E. Highly cited authors (commentary). Scientist. 2002; 16(7): 10-1.

26. Campanario M. Consolation for the scientist: sometimes it is hard to publish papers that are later highly cited. Soc Stud Sci. 1993; 23(2): 342-62. doi: 10.1177/030631293023002005.

27. Rushton JP, Murray HG, Paunonen SV. Personality, research creativity, and teaching effectiveness in university professors. Scientometrics. 1983; 5(2): 93-116. doi: 10.1007/BF02072856.

28. Hogan JD, Hedgepeth R. Journal quality: the issue of diversity. Am Psychol. 1983; 38(8): 961-2. doi: 10.1037/0003-066X.38.8.961.

29. Price DD. A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inf Sci. 1976; 27(5): 292-306. doi: 10.1002/asi.4630270505.

30. Lovaglia MJ. Status characteristics of journal articles for editor's decisions and citations. The Society for Social Studies of Science Annual Meeting, November; 1989. p. 15-8.

31. Glänzel W, Debackere K, Thijs B, Schubert A. A concise review on the role of author self-citations in information science, bibliometrics and science policy. Scientometrics. 2006; 67(2): 263-77. doi: 10.1007/s11192-006-0098-9.

32. Jones BF, Wuchty S, Uzzi B. Multi-university research teams: shifting impact, geography, and stratification in science. Science. 2008; 322(5905): 1259-62. doi: 10.1126/science.1158357, PMID 18845711.

33. Beyer JM. Editorial policies and practices among leading journals in four scientific fields. Sociol Q. 1978; 19(1): 68-88. doi: 10.1111/j.1533-8525.1978.tb02172.x.

34. Hargens LL. Using the literature: reference networks, reference contexts, and the social structure of scholarship. Am Sociol Rev. 2000; 65(6): 846-65. doi: 10.1177/000312240006500603.

35. Shadish WR, Tolliver D, Gray M, Sen Gupta SK. Author judgements about works they cite: three studies from psychology journals. Soc Stud Sci. 1995; 25(3): 477-98. doi: 10.1177/030631295025003003.

36. Funk RJ, Owen-Smith J. A dynamic network measure of technological change. Manag Sci. 2017; 63(3): 791-817. doi: 10.1287/mnsc.2015.2366.

37. Lee YN, Walsh JP, Wang J. Creativity in scientific teams: unpacking novelty and impact. Res Policy. 2015; 44(3): 684-97. doi: 10.1016/j.respol.2014.10.007.

38. Uzzi B, Mukherjee S, Stringer M, Jones B. Atypical combinations and scientific impact. Science. 2013; 342(6157): 468-72. doi: 10.1126/science.1240474, PMID 24159044.

39. Wang J, Veugelers R, Stephan P. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Res Policy. 2017; 46(8): 1416-36. doi: 10.1016/j.respol.2017.06.006.

40. Leahey E. Not by productivity alone: how visibility and specialization contribute to academic earnings. Am Sociol Rev. 2007; 72(4): 533-61. doi: 10.1177/000312240707200403.

41. Leahey E, Keith B, Crockett J. Specialization and promotion in an academic discipline. Res Soc Stratification Mob. 2010; 28(2): 135-55. doi: 10.1016/j.rssm.2009.12.001.

42. Heiberger H, Wieczorek J, Wäckerle J. Scientific competition in the attention economy: focus as a strategic resource. Minerva. 2021; 59(3): 255-76.

43. Whitley R. The intellectual and social organization of the sciences. 2nd ed. Oxford University Press; 2000. doi: 10.1093/oso/9780199240531.001.0001.

44. Abbott A. The chaos of disciplines. University of Chicago Press; 2001.

45. Cañibano C, Otamendi FJ, Solís F. Specialisation and career paths in science: the case of researchers in the Spanish CSIC. Scientometrics. 2019; 118(2): 709-18.

46. Hackett EJ. Essential tensions: identity, control, and risk in research. Soc Stud Sci. 2005; 35(5): 787-826. doi: 10.1177/0306312705056045.

47. Fischer T, Leidinger J. Testing patent value indicators on directly observed patent value-an empirical analysis of Ocean Tomo patent auctions. Res Policy. 2014; 43(3): 519-29. doi: 10.1016/j.respol.2013.07.013.

48. Shi F, Evans J. Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. Nat Commun. 2023; 14(1): 1641. doi: 10.1038/s41467-023-36741-4, PMID 36964138.

49. Bornmann L, de Moya Anegón F, Leydesdorff L. Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. PLOS One. 2010; 5(10): e13327. doi: 10.1371/journal.pone.0013327, PMID 20967252.

50. Veugelers R, Wang J. Scientific novelty and technological impact. Res Policy. 2019; 48(6): 1362-72. doi: 10.1016/j.respol.2019.01.019.

51. Ruan X, Ao W, Lyu D, Cheng Y, Li J. Effect of the topic-combination novelty on the disruption and impact of scientific articles: evidence from PubMed. J Inf Sci. 2023: 01655515. doi: 10.1177/01655515231161133.

52. Foster JG, Rzhetsky A, Evans JA. Tradition and innovation in scientists' research strategies. Am Sociol Rev. 2015; 80(5): 875-908. doi: 10.1177/0003122415601618.

53. Heiberger RH, Munoz-Najar Galvez S, McFarland DA. Facets of specialization and its relation to career success: an analysis of US Sociology, 1980 to 2015. Am Sociol Rev. 2021; 86(6): 1164-92. doi: 10.1177/00031224211056267.