State of the Art of Novelty Indicators

Sotaro Shibayama *, Zhao Wu , Deyun Yin , Kazuki Yokota , Pierre Pelletier
The University of Tokyo

04-Mar-2026
DZHW Seminar

Motivation

- Novelty (originality) is a critical concept in science (Barber, 1952; Bourdieu, 1975; Kuhn, 1970; Merton, 1973; Polanyi, 1962).
- Many indicators of novelty exist, but their validity has been elusive.
  - They are largely driven by convenient operationalizations (new word, etc.).
  - Recent studies tend to include some validation exercises, but insufficient (based on weak proxies, etc.).
- Despite the limitation, more novelty indicators are developed and made publicly available (Arts et al., 2025; Lin et al., 2023), which could mislead scholarly as well as policy debates.

Aim

- To test novelty indicators against self-reported (survey) measures
- To understand the limitations of the indicators and provide suggestions for the future development

Key findings

Previous limitation 1: Multifaceted nature of novelty has been overlooked.

→ While most indicators detect some types of novelty, a single indicator does not sufficiently capture all types of novelty.

Previous limitation 2: Novelty may mean different things across disciplines, but such disciplinary differences have been largely overlooked.

→ Indicators' performance varies across disciplines. Some indicators demonstrate consistent validity across many disciplines, while others show validity only in limited disciplines.

Outline

1. Theoretical background
2. Existing indicators and their limitations
3. Study design – questionnaire and bibliometrics
4. Result
5. Discussion

1. Theoretical background

- Novelty constitutes a core value in science. – Highly abstract argument
  - Novelty (originality) as a scientific norm (Hagstrom, 1974; Merton, 1973)
  - Novel discoveries make impact on science and innovation (Veugelers and Wang, 2019).
  - Novelty serves as a key criterion for funding allocation, employment, and scientific awards (Dasgupta and David, 1994; Merton, 1973; Stephan, 1996; Storer, 1966).

- Efforts to dissect the novelty concept
  - Dirk (1999): Questionnaire in life sciences
    - Originality in a hypothesis; originality in a method; originality in a result
  - Guetzkow et al. (2004): Interviews of grant referees in HASS
    - 7 categories of originality (approach, topic, theory, method, data, result, and understudied area)
  - Barlösius (2019): Reviewing funding proposals in STEM
    - 3 modes of claiming the originality of funding proposals: doing something (1) for the first time, (2) partly differently from a customary manner, and (3) in a totally different manner.

- Learnings from the sociology of science literature
  - Novelty is a multifaceted concept – research can be novel in various aspects.
  - Typical novelty aspects may differ between disciplines. Not all aspects of novelty are equally valued – scientists may appreciate a certain type of novelty than other types.
  - The novelty concept itself is still elusive (, which this study does not address…)

2. Existing indicators

- A new word/keyword
  - The vintage of MeSH keyword (Azoulay et al., 2011)
  - A new word used for the first time in history (Arts et al., 2025)

- Recombination
  - A rare combination of journals (Lee et al., 2015; Uzzi et al., 2013; Wang et al., 2017)
  - A rare combination of references (Dahlin and Behrens, 2005; Matsumoto et al., 2021; Trapido, 2015)
  - A new combination of words (Arts et al., 2025; Boudreau et al., 2016)

- Based on document embeddings
  - Isolated paper (Yin et al., 2023: Arts et al., 2025)
  - Distant combination of references (Shibayama et al., 2021)

- More recent development
  - Use of full text (Luo et al., 2022; Wang, 2024)
  - Combination of research question and method (identifying different parts of a paper) (Luo et al., 2022)
  - Directly tasking AI to assess novelty (Zhang and Chen, 2024)

2. Existing indicators – categorization

1. Ex-ante vs. ex-post
   - Ex-ante: based on information available at the time of publication (most novelty indicators)
   - Ex-post: require information generated after publication (e.g., disruptiveness)

2. Recombination of objects vs. an object itself
   - Recombination: based on newness/rarity of combination of multiple objects (e.g., journal recombination, a new pair of words)
   - Object: based on the newness of a single object (e.g., a new word, isolated paper)

3. Unit of objects
   - Word, sentence, reference, journal, etc.

4. Citation-based vs. text-based
   - Citation: draw on citation network data (e.g., journal combination, reference combination)
   - Text: draw on text information such as title and abstract (a new word, ones based on document-embedding)

2. Existing indicators – categorization

Ex-post vs. Ex-ante	Recombination vs. Object	Unit of object	Data type	Indicator	Description	Source
Ex-ante	Recombination	Journal	Citation	*JournalComb*	A rare combination of cited journals	Lee, et al. (2015)
		Paper (Reference)	Citation & Text vector	*RefComb*	A distant combination of cited references	Shibayama, et al. (2021)
		Word	Text	*WordComb*	A new combination of phrases	Arts, et al. (2025)
	Object	Paper	Text vector	*DocDistant*	Distance from extant papers	Yin, et al. (2023)
		Word	Text	*WordNew*	A new phrase	Arts, et al. (2025)
Ex-post	N/A	Paper	Citation	*Disruption*	Disruptiveness indicator	Wu, et al. (2019)
		Word	Text	*WordReuse*	The number of times a new phrase is reused	Arts, et al. (2025)

2. Existing indicators – limited validation

Weak proxy of novelty

- Comparison of Nobel laureates' papers and non-laureates' papers (Arts et al., 2025; Wu et al., 2019)
- Comparison of review articles and non-review articles (Arts et al., 2025; Teplitskiy et al., 2022)

→ Correspondence to the novelty concept is ambiguous

→ Multifaceted nature of novelty is ignored.

Tested in a narrow domain

- Faculty Opinion (a.k.a. F1000, H1, Faculty Review): provides experts' assessments, but only in biomedicine (Bornmann et al., 2019; Teplitskiy et al., 2022; Wang et al., 2024).
- Peer review data: only in a particular conference (Wang et al., 2024)

→ Sensitivity to disciplinary difference is overlooked.

* Some studies carried out validation analyses including multiple disciplines but do not disclose potential differences across disciplines.

3. Study design – sample

- Sample = 2,226 papers published by the survey respondents
- Online questionnaire (1) in 2021-2022 and (2) in 2024
  - 2,226 responses in total (the response rate = 9.3%)
- Sampling criteria
  - Identified corresponding authors whose email addresses are available in Web of Science (WoS).
  - Listed all English-written papers ("article", "letter", or "proceedings paper")
  - Selected the latest paper published up to 2018 (round 1) or 2019 (round 2) for each author.
  - Randomly sampled the respondents with scientific fields stratified

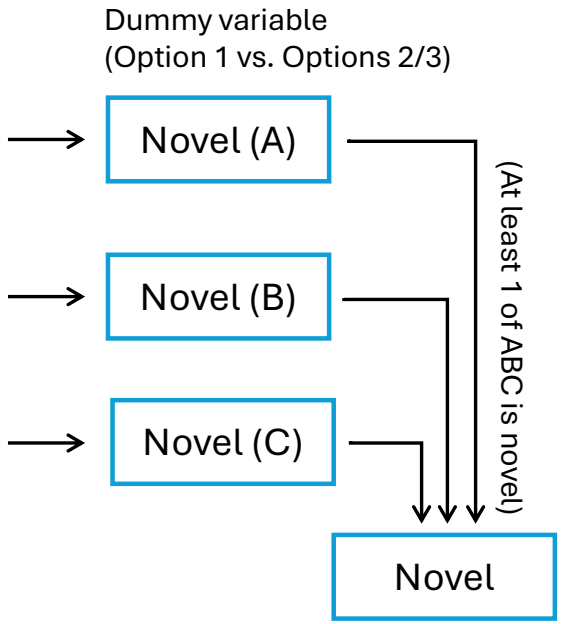
Note: The 2 surveys asked partially different questions, so the sample size of each analysis can vary.

3. Study design – survey question (1)

The following questions are about your paper:
"Smith et al. (2020) XXXXXXXX"

How do you describe the **contribution** of your paper **compared to previous papers**?
Please assess your paper in the following aspects (A, B, and C) **respectively**.

	[1] **Novel (not previously reported)**	[2] Improvement on previous papers	[3] Not applicable (previously reported or irrelevant)
(A) In **creating/developing** something (e.g., methods, materials, data, chemicals)	☐	☐	☐
(B) In **discovering/identifying** something (e.g., phenomena, substances, molecules)	☐	☐	☐
(C) In **understanding/explaining** something (e.g., mechanisms, theories)	☐	☐	☐



3. Study design – survey question (2)

Considering how your paper was typically cited, how unique information do you think you paper offered to them?

[1] Not unique. Easily replaceable.
[2] Not unique. But, offers improvement over others
[3] Unique. But, there may be work-around.
[4] Unique and essential.

→ Irreplaceable

Dummy variable
(Options 3/4 vs. Options 1/2)

3. Study design – survey question (3)

At the time of publication, how was your paper positioned in the existing scholarly debate?

[1] In line with the mainstream argument
[2] In line with one of the multiple competing arguments
[3] Against the mainstream argument
[4] No active debate existed in the topic.

→ NewTopic

Dummy variable
(Option 4 vs. Options 1/2/3)

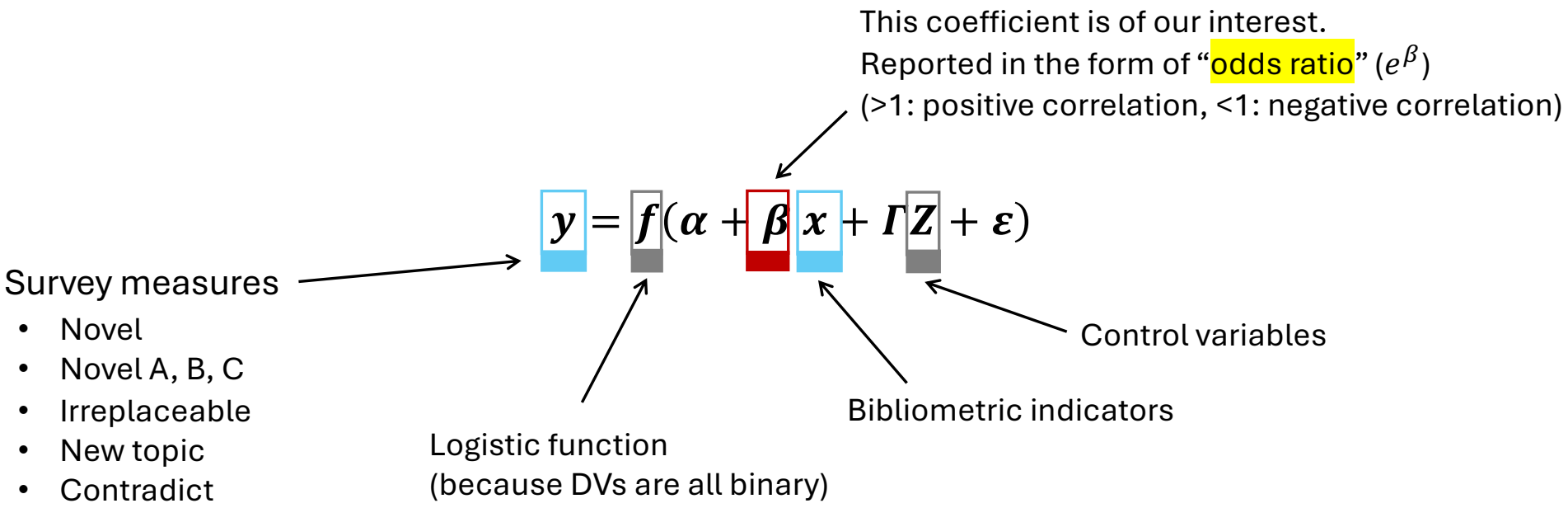
→ Contradict

Dummy variable
(Option 3 vs. Options 1/2)

3. Study design – bibliometric indicators

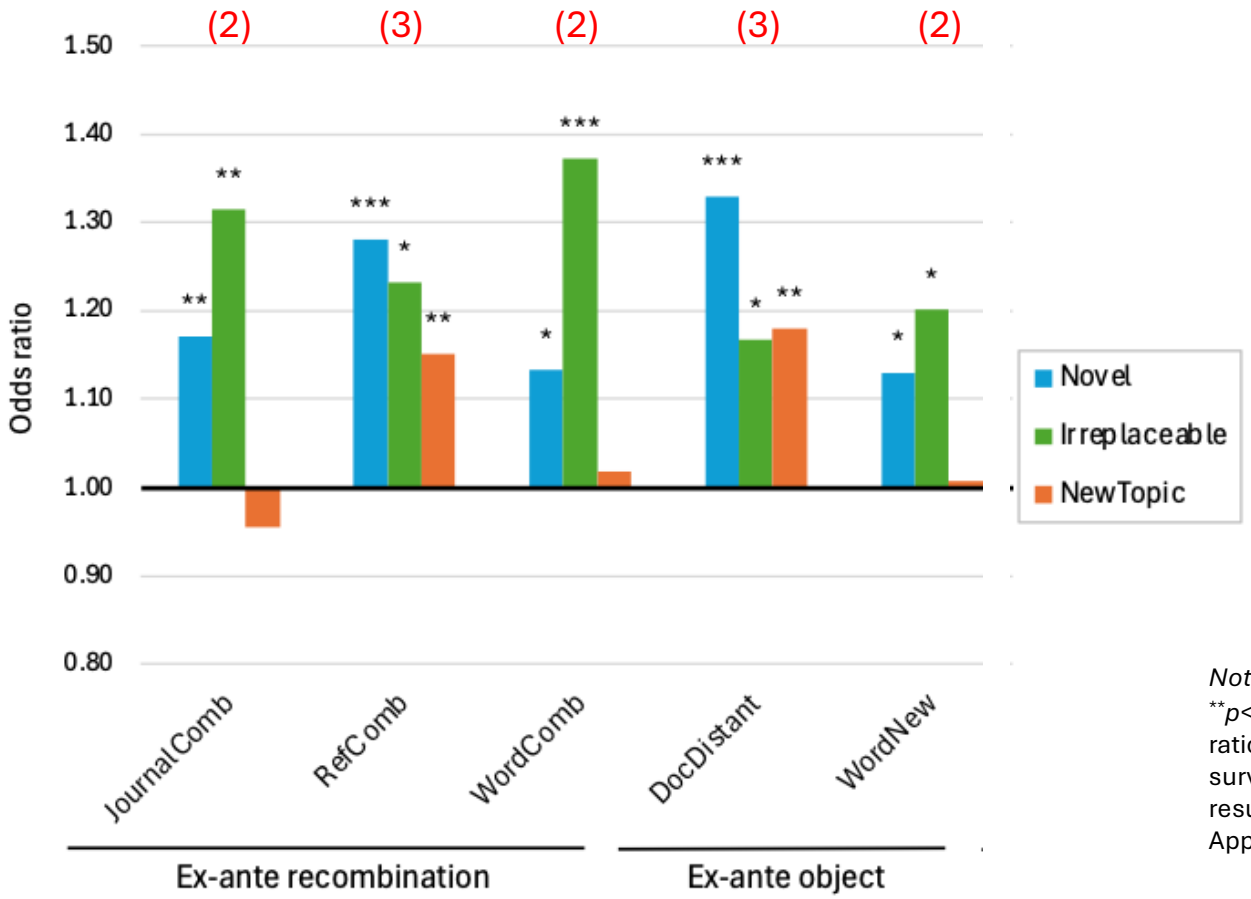
Ex-post vs. Ex-ante	Recombination vs. Object	Unit of object	Data type	Indicator	Description	Source
Ex-ante	Recombination	Journal	Citation	*JournalComb*	A rare combination of cited journals	Lee, et al. (2015)
		Paper (Reference)	Citation & Text vector	*RefComb*	A rare combination of cited references	Shibayama, et al. (2021)
		Word	Text	*WordComb*	A new combination of phrases	Arts, et al. (2025)
	Object	Paper	Text vector	*DocDistant*	Distance from extant papers	Yin, et al. (2023)
		Word	Text	*WordNew*	A new phrase	Arts, et al. (2025)
Ex-post	N/A	Paper	Citation	*Disruption*	Disruptiveness indicator	Wu, et al. (2019)
		Word	Text	*WordReuse*	The number of times a new phrase is reused	Arts, et al. (2025)

3. Study design – Econometrics



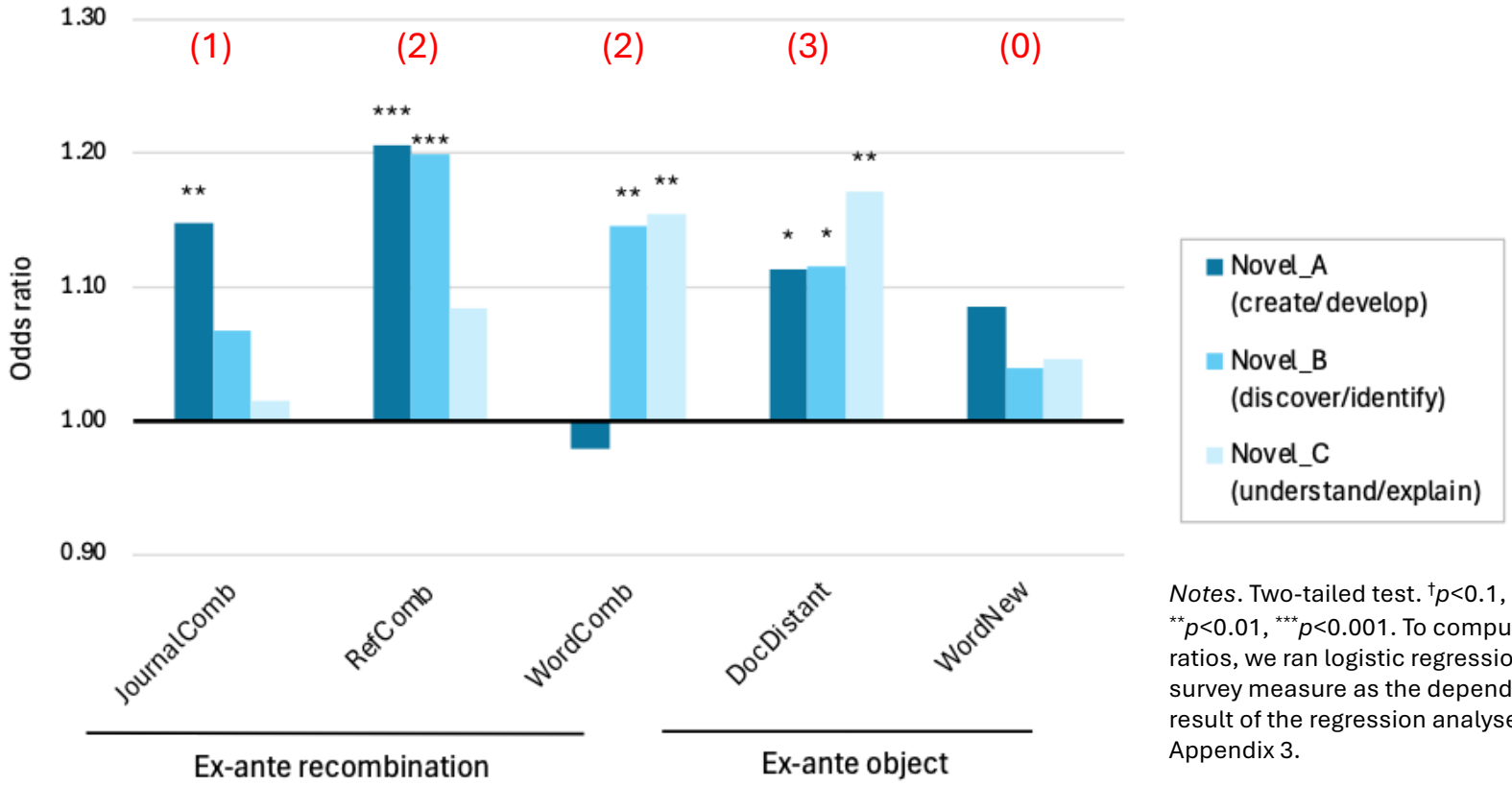
Note: Bibliometric indicators are often influenced by scientific fields, for example, because citation patterns or text expressions are characteristic to certain disciplines. To address this issue, we standardised all bibliometric indicators by linear transformation so that the mean of the standardised indicators becomes 0 and the standard deviation becomes 1 in respective fields (Agricultural science, Medicine, Biology, Chemistry, Material science, Engineering, and Physics). Our main analyses do not include any control variable, which makes the most conservative validation, given that the indicators are likely to be used without such control variables.

4. Result – (1) While most indicators detect some aspects of novelty, a single indicator may not sufficiently capture all forms of novelty.



Notes. Two-tailed test. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. To compute the odds ratios, we ran logistic regressions for each survey measure as the dependent variable. The result of the regression analyses is presented in Appendix 3.

4. Result – (1) While most indicators detect some aspects of novelty, a single indicator may not sufficiently capture all forms of novelty.



Notes. Two-tailed test. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. To compute the odds ratios, we ran logistic regressions for each survey measure as the dependent variable. The result of the regression analyses is presented in Appendix 3.

4. Result – (2) Some indicators demonstrate consistent correlations across disciplines while others show correlations only in limited disciplines.

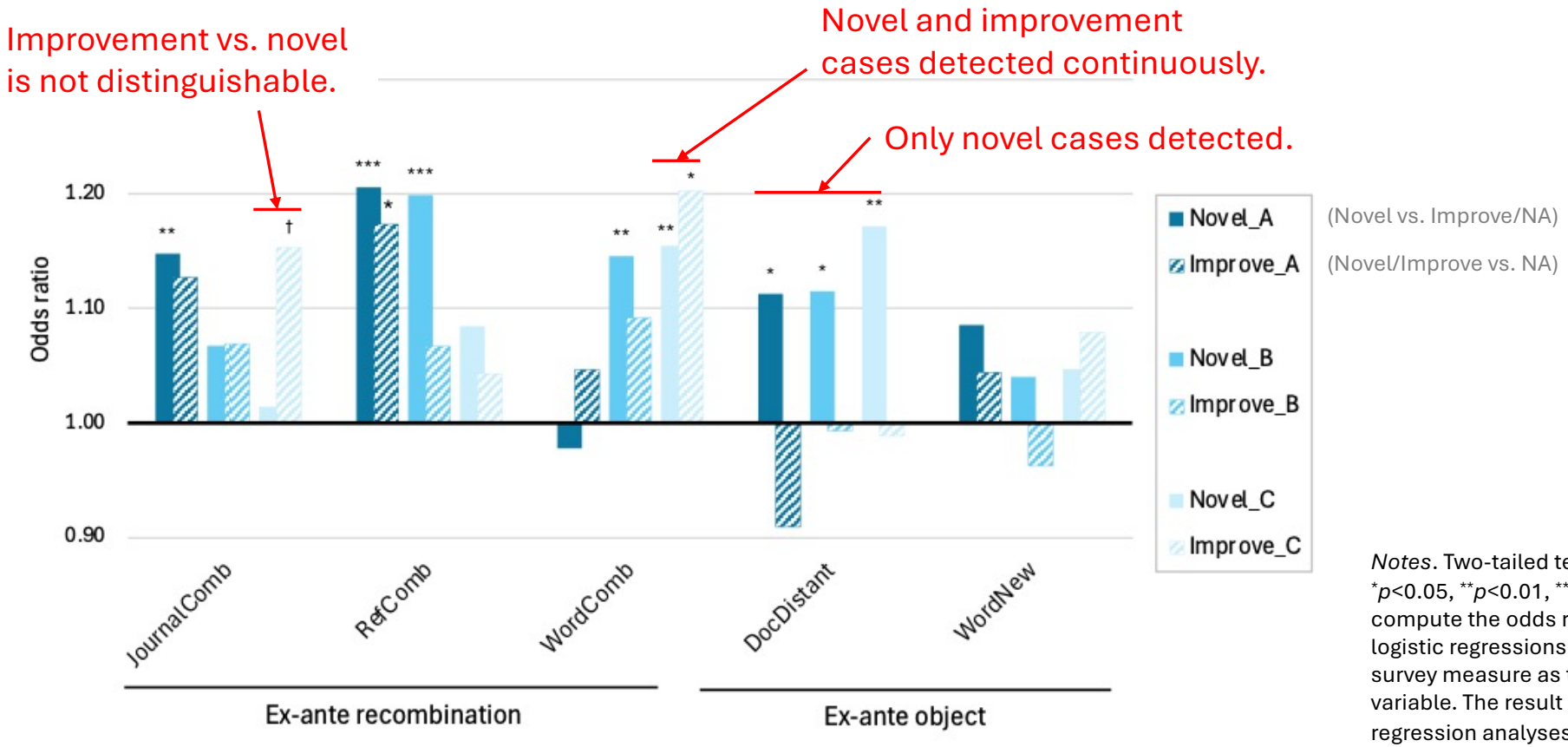
			JournalComb	RefComb	WordComb	DocDistant	WordNew	Disruption	WordReuse
Novel	Agriculture	(205)	0.91	1.13	1.10	**1.37**	1.09	1.35	1.07
	Medicine	(255)	**1.70**	**1.82**	**1.34**	**1.55**	**1.30**	1.51	**1.28**
	Biology	(306)	1.18	1.14	**1.28**	**1.36**	**1.29**	0.24	1.21
	Chemistry	(346)	0.89	**1.26**	0.98	**1.48**	1.25	0.56	1.26
	Material	(254)	**1.43**	1.31	0.93	1.07	1.09	0.37	1.10
	Engineering	(445)	1.11	**1.20**	0.97	**1.33**	0.95	0.99	0.97
	Physics	(269)	**1.56**	1.24	1.15	1.19	1.03	0.63	1.00
Irreplaceable	Agriculture	(94)	1.44	**1.63**	1.44	1.25	**1.33**	2.42	1.38
	Medicine	(185)	**1.92**	**1.35**	1.10	1.06	1.35	0.67	1.35
	Biology	(222)	**2.19**	**1.67**	**1.70**	1.22	**1.53**	0.10	**1.46**
	Chemistry	(241)	1.27	**1.42**	**1.33**	1.23	1.17	0.27	1.24
	Material	(133)	1.65	1.22	**1.52**	1.35	1.00	0.62	1.01
	Engineering	(278)	1.17	1.15	**1.30**	**1.38**	1.02	0.95	1.04
	Physics	(159)	1.28	0.89	**1.53**	0.93	1.24	0.56	1.14
New Topic	Agriculture	(226)	0.57	1.12	**1.42**	1.13	1.28	1.13	1.34
	Medicine	(305)	0.88	1.13	1.04	1.24	1.11	1.31	1.09
	Biology	(368)	1.15	**1.25**	1.04	**1.26**	0.96	1.21	0.93
	Chemistry	(407)	1.10	1.21	0.87	1.04	0.94	1.14	0.94
	Material	(290)	1.04	**1.33**	0.88	**1.30**	1.03	0.94	1.08
	Engineering	(503)	1.15	**1.23**	0.92	**1.27**	0.99	1.14	0.99
	Physics	(320)	**1.27**	**1.38**	0.86	**1.30**	0.86	0.80	0.78

High consistency

Medium consistency

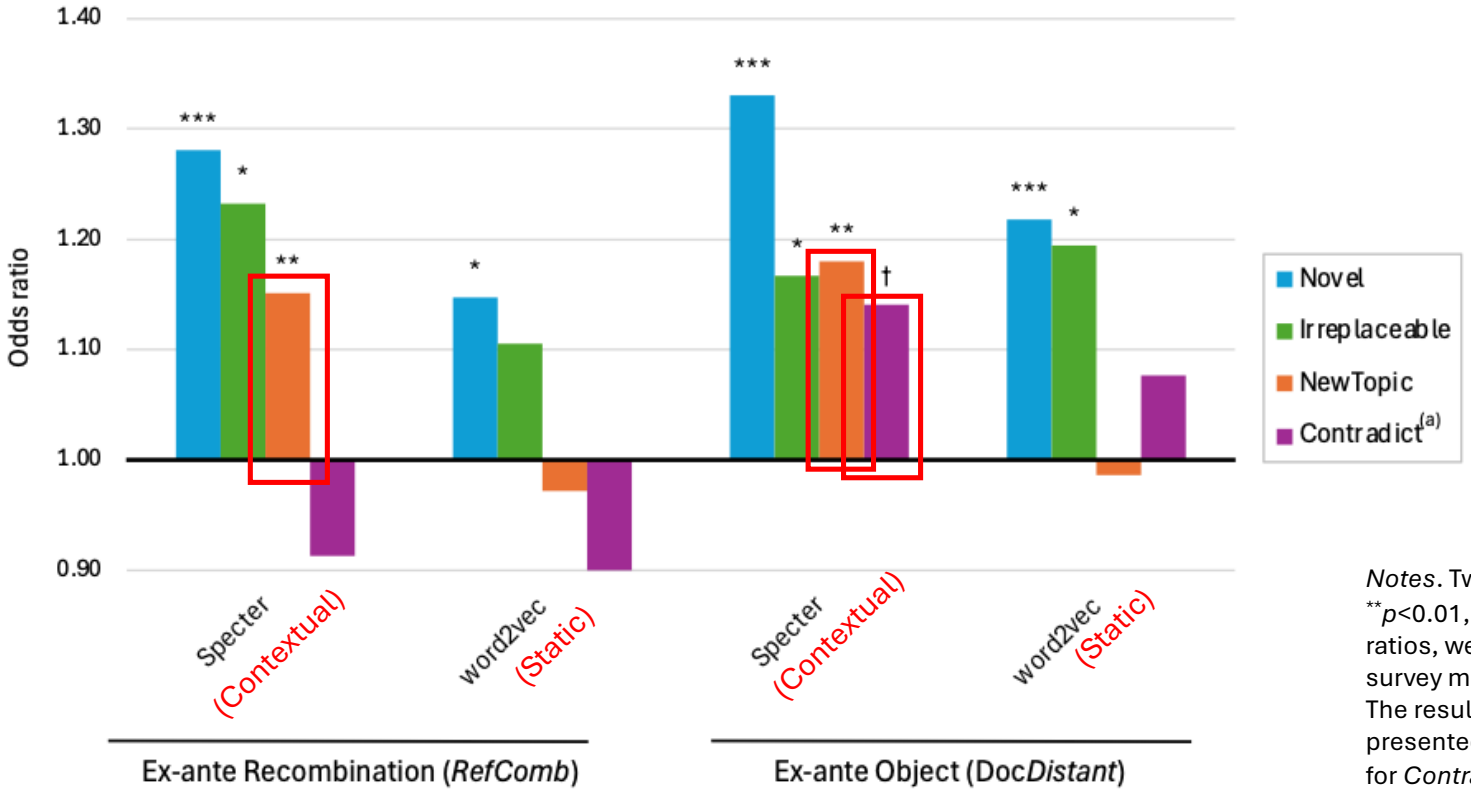
Notes. We ran a logistic regression for each pair of bibliometric indicators as the independent variable and survey measures as the dependent variable in each field and displayed the odds ratios. The cell colour reflects the magnitude of the odds ratios, and the bold type indicates statistical significance at the 10% level. In the parentheses are the sample size (the average across all bibliometric indicators). As the statistical significance is partly influenced by the difference in sample sizes, the interpretation of the results should also consider the magnitude of odds ratios (cell colours). The result of the regression analyses is presented in Appendix 3.

4. Result – (3) Some indicators capture both high and low levels of novelty along a continuum, others detect specifically high-level novelty, and yet others detect the presence of newness without differentiating its degree



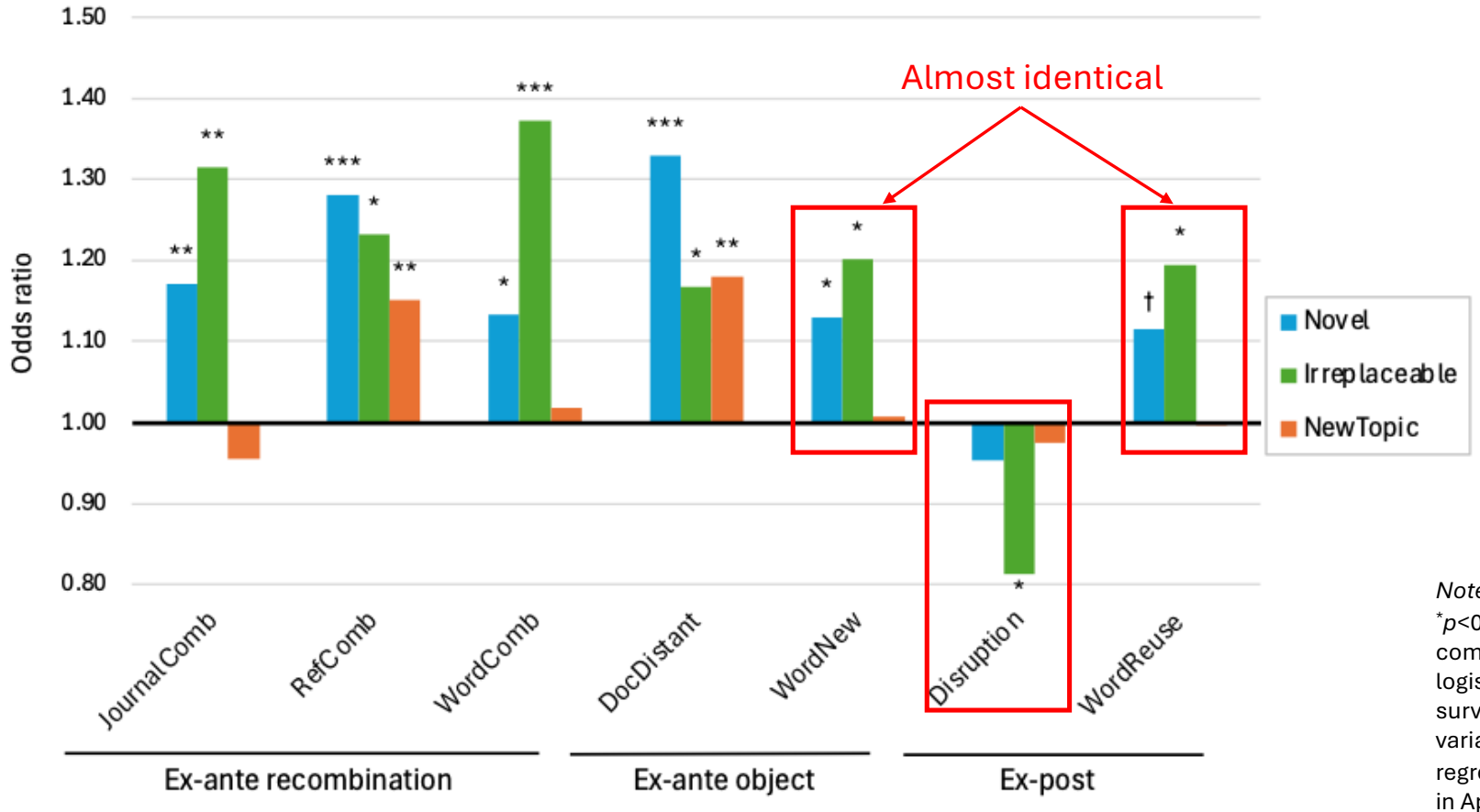
Notes. Two-tailed test. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. To compute the odds ratios, we ran logistic regressions for each survey measure as the dependent variable. The result of the regression analyses is presented in Appendix 3.

4. Result – (4) Comparing static and contextual language models, we find that indicators based on the latter outperforms those on the former.



Notes. Two-tailed test. †$p<0.1$, *$p<0.05$, **$p<0.01$, ***$p<0.001$. To compute the odds ratios, we ran logistic regressions for each survey measure as the dependent variable. The result of the regression analyses is presented in Appendix 3. [a]: The regressions for *Contradict* include a few control variables as they fail to find significant correlations otherwise.

4. Result – (5) Ex-post indicators, which require post-publication data (e.g., forward citation), offer no clear advantages over ex-ante indicators in detecting novelty.



Notes. Two-tailed test. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. To compute the odds ratios, we ran logistic regressions for each survey measure as the dependent variable. The result of the regression analyses is presented in Appendix 3.

4. Result – variants of bibliometric indicators

Independent variables (Bibliometric indicators)		Dependent variable (Survey measure)								
		Novel			Irreplaceable			NewTopic		
Indicator	Variant	O.R.	p	N	O.R.	p	N	O.R.	p	N
JournalComb	Lee, et al. (2015)	**1.171**	**(0.004)**	**1583**	**1.315**	**(0.001)**	**959**	0.955	(0.374)	1828
	Uzzi, et al.(2013) [a]	1.063	(0.240)	1528	**1.331**	**(0.013)**	**936**	0.975	(0.625)	1772
WordNew	Noun Phrase	**1.129**	**(0.042)**	**1586**	**1.201**	**(0.016)**	**964**	1.008	(0.886)	1835
	Word [b]	0.988	(0.828)	1586	1.048	(0.514)	964	0.977	(0.668)	1835
WordComb	Noun Phrase	**1.134**	**(0.024)**	**1586**	**1.372**	**(0.000)**	**964**	1.018	(0.733)	1835
	Word [b]	**1.125**	**(0.039)**	**1586**	**1.268**	**(0.001)**	**964**	0.987	(0.803)	1835
WordReuse	Noun Phrase	**1.114**	**(0.069)**	**1586**	**1.195**	**(0.020)**	**964**	0.996	(0.942)	1835
	Word [b]	0.975	(0.658)	1586	1.061	(0.419)	964	0.986	(0.788)	1835
DocDistant	Yin, et al. (2023)	**1.330**	**(0.000)**	**1601**	**1.167**	**(0.046)**	**986**	**1.180**	**(0.002)**	**1858**
	Arts, et al. (2025) [c]	1.073	(0.212)	1586	1.128	(0.150)	964	**1.118**	**(0.033)**	**1835**

Note: Yellow highlights show the tested indicators, which shows advantage over the alternatives.

4. Result – Nobel laureates vs. non-laureates

Independent variables (Bibliometric indicators)	Dependent variable: Nobel award		
	O.R.	p	N
JournalComb	1.389	(0.022)	630
RefComb	2.233	(0.000)	536
WordComb	5.604	(0.000)	1136
DocDistant	1.362	(0.030)	686
WordNew	2.520	(0.000)	1136
Disruption	6.719	(0.000)	708
WordReuse	2.719	(0.000)	1136

Notes. We downloaded Nobel laureates' publication metadata constructed by (Li et al., 2019). The dataset includes 730 publications between 1887 and 2010 identified by OpenAlex (OA) paper ID. We then prepared a control sample for each one of these publications by finding papers published in the same year, in the same journal, and in the same OA subfield (if no paper was found in the same subfield, we relaxed the condition to find papers in the same OA field). As no paper that satisfies these conditions was found for some of Nobel laureates' papers, we ended up with 684 pairs of laureates' and non-laureates' publications. We then linked these publications with bibliometric indicators – we downloaded word-based indicators and disruption from open data sources (https://zenodo.org/records/13869486 and https://github.com/kellogg-cssi/SciSciNet) and computed other indicators based on OA. The sample size for the final analysis (*N*) differs between indicators, either because the corresponding publication is not included in the data sources or because OA does not provide sufficient information to compute the indicators. Then, we ran logistic regressions, in which the dependent variable is the dummy variable coded 1 for Nobel laureates' publications and 0 for non-laureates' (*Nobel award*) and each bibliometric indicator as the independent variable. To make direct comparison between each pair, we incorporated the fixed effects of each pair. For each regression, the table reports the odds ratio (O.R.), p-value (p), and sample size (N).

5. Discussion – summary

- Correspondence to novelty concepts
  - While most indicators detect some aspects of novelty, a single indicator may not sufficiently capture all forms of novelty.
  - Some indicators capture both high and low levels of novelty along a continuum, others detect specifically high-level novelty, and yet others detect the presence of newness without differentiating its degrees.

- Disciplinary sensitivity
  - Some indicators demonstrate consistent correlations across many disciplines while others show correlations only in limited disciplines.

- Misc.
  - Comparing static and contextual language models, we find that indicators based on the latter outperforms those on the former.
  - Ex-post indicators do not offer clear advantages over ex-ante indicators in detecting novelty.

5. Discussion – direction

- For indicator users
  - Be aware of the multifaceted nature of novelty and select appropriate ones based on their needs.
  - Decide on indicators suitable for disciplines of their interest.
  - Advisable not to rely solely on a single indicator, but to test multiple indicators.
  - Correlations between the survey measures and the bibliometric indicators are not always strong. This implies that the indicators may exhibit non-negligible errors, and they should be used in an aggregated manner to mitigate such errors.

- For indicator developers
  - Develop indicators for the aspects of novelty that are poorly addressed by existing indicators (e.g., *Contradict*)
  - Develop an indicator that captures a broader range of novelty aspects
  - Develop indicators applicable to various disciplines
  - Elaborate the link between theories and indicators