

Bias against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators

Jian Wang ^{a, b}, Reinhilde Veugelers ^{a, c, d, *} and Paula Stephan ^{e, f}

^a Department of Managerial Economics, Strategy and Innovation (MSI) and Center for R&D Monitoring (ECONOM), KU Leuven, Leuven, Belgium.

^b Labor and Worklife Program, Harvard Law School, Cambridge, MA, United States.

^c Bruegel, Brussels, Belgium.

^d Center for Economic Policy Research (CEPR), London, United Kingdom.

^e Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, United States.

^f National Bureau of Economic Research (NBER), Cambridge, MA, United States.

February 27, 2017

* Corresponding author at: Naamsestraat 69, 3000 Leuven, Belgium. E-mail address: reinhilde.veugelers@kuleuven.be (R. Veugelers).

Acknowledgements: Earlier versions of this paper were presented at the Workshop on the Organization, Economics and Policy of Scientific Research, Turin; Institute for Research Information and Quality Assurance, Berlin; Max Planck Institute for Innovation and Competition, Munich; TIES seminar at the MIT Sloan School, Cambridge, Economics of Science and Engineering Workshop at Harvard University, Cambridge; OECD Blue Sky Forum, Ghent; and REER, Atlanta. The authors thank the reviewers, the editor, and seminar participants, in particular Pierre Azoulay, Christian Catalini, Paul David, Lee Fleming, Richard Freeman, Alfonso Gambardella, Dietmar Harhoff, Diana Hicks, Sybille Hinze, Stefan Hornbostel, Jacques Mairesse, Fabio Montobbio, Henry Sauermann, Daniel Sirtes, Scott Stern, Mark Veugelers, and John Walsh for helpful and encouraging comments. Financial support from KU Leuven (GOA/12/003) and the Research Foundation - Flanders (FWO, G.0825.12) is gratefully acknowledged. J. Wang also gratefully acknowledges a postdoctoral fellowship from FWO. Publication data are sourced from Web of Science Core Collection.

Jian Wang, Reinhilde Veugelers & Paula Stephan (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416-1436. <http://dx.doi.org/10.1016/j.respol.2017.06.006>

© 2017 Elsevier B.V. All rights reserved.

ABSTRACT

Research which explores uncharted waters has a high potential for major impact but also carries a higher uncertainty of having impact. Such explorative research is often described as taking a *novel* approach. This study examines the complex relationship between pursuing a novel approach and impact. Viewing scientific research as a combinatorial process, we measure novelty in science by examining whether a published paper makes first-time-ever combinations of referenced journals, taking into account the difficulty of making such combinations. We apply this newly developed measure of novelty to all Web of Science research articles published in 2001 across all scientific disciplines. We find that highly novel papers, defined to be those that make more (distant) new combinations, deliver high gains to science: they are more likely to be a top 1% highly cited paper in the long run, to inspire follow-on highly cited research, and to be cited in a broader set of disciplines and in disciplines that are more distant from their “home” field. At the same time, novel research is also more risky, reflected by a higher variance in its citation performance. We also find strong evidence of delayed recognition of novel papers as novel papers are less likely to be top cited when using short time-windows. In addition, we find that novel research is significantly more highly cited in “foreign” fields but not in their “home” field. Finally, novel papers are published in journals with a lower Impact Factor, compared with non-novel papers, *ceteris paribus*. These findings suggest that science policy, in particular funding decisions which rely on bibliometric indicators based on short-term citation counts and Journal Impact Factors, may be biased against “high risk/high gain” novel research. The findings also caution against a mono-disciplinary approach in peer review to assess the true value of novel research.

Keywords: novelty, breakthrough research, bibliometrics, evaluation, impact

JEL: I23,O31,O33,O38

1. Introduction

Scientific breakthroughs advance the knowledge frontier. Research underpinning breakthroughs often is driven by novel approaches. While research that takes a novel approach has a higher potential for major impact, it also faces a higher level of uncertainty of impact. In addition, it may take longer for novel research to have a major impact, displaying a profile of *scientific prematurity* (Stent, 1972), *delayed recognition* (Garfield, 1980), or that of a *sleeping beauty* (Van Raan, 2004), either because of resistance from incumbent scientific paradigms (Kuhn, 1962; Merton, 1973; Planck, 1950) or because of the longer time required to recognize and incorporate the findings of novel research into follow-on research (Garfield, 1980; Wyatt, 1975). The “high risk/high gain” nature of novel research makes it particularly appropriate for public support (Arrow, 1962). Delayed recognition may, however, lead novel research to be undervalued in research evaluations which rely on indicators based on short term citation windows.

Any bias in commonly used bibliometric indicators against novel research, to the extent it exists, is of concern given the increased reliance funding agencies and hiring institutions place on readily available bibliometric information to aid in decision making and performance evaluation (Butler, 2003; Hicks, 2012; Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015; Martin, 2016; Monastersky, 2005). Such heavy reliance may explain in part the perception that funding agencies and their expert panels are increasingly risk-averse and the charge that competitive selection procedures encourage relatively safe projects, which exploit existing knowledge, at the expense of novel projects that explore untested approaches (Alberts, 2010; Azoulay, Graff Zivin, & Manso, 2011; Kolata, 2009; NPR, 2013; Petsko, 2012; Walsh, 2013).

The goal of this paper is to develop a measure of novel research and compare the citation profile of novel research with that of non-novel research, as well as the Impact Factor of the journals in which novel research is published. We are particularly interested in whether the impact profile of novel research matches the “high risk/high gain” profile associated with breakthrough research and which commonly used bibliometric measures would be biased against novel research. To this end, we define research that draws on new combinations of knowledge components as novel and develop an *ex ante* measure of combinatorial novelty at the paper level, where novelty is operationalized as making new combinations in referenced journals. Utilizing this newly-minted measure of novelty, we explore the complex relationship between novelty and citation impact,

using the life-time citation trajectories of research articles across all scientific disciplines published in 2001 and indexed in the Web of Science (WoS), as well as the profile of papers citing them.

We find novel papers to have a larger variance in their citation distribution and be more likely to populate both the tail of high impact and the tail of low impact, reflecting their “high risk” profile. At the same time, novel papers also display a “high gain” characteristic: they have a much higher chance of being a top cited paper in the long run, a higher likelihood of stimulating follow-on top cited research, and a broader impact transcending disciplinary boundaries and reaching more distant scientific fields. We further scrutinize the impact profile of novel research and uncover intriguing characteristics associated with novelty. First, we distinguish between impact in “home” and “foreign” fields and find that, compared with non-novel papers, novel papers are significantly more likely to be highly cited in foreign fields but not in their home field. Second, an examination of time dynamics in the citation accumulation process reveals delayed recognition for novel research. Specifically, although novel papers are highly cited in the long run, they are less likely to be top cited in the short run. We also find that novel papers are less likely to be published in high Impact Factor journals. These findings suggest that over-reliance on Journal Impact Factor and citation counts using short citation time-windows, may bias against novel research.

2. Combinatorial novelty in science

Scientific discovery can be viewed as a form of human problem solving (Klahr & Simon, 1999; Simon, 1966; Simon, Langley, & Bradshaw, 1981), the process for which involves a combinatorial aspect, such as integrating different perspectives for defining the problem space and assembling various methods and tools for solving the problem within the problem space. In this respect, the creation of new scientific knowledge builds on combining existing pieces of knowledge. Some of these existing knowledge pieces are embedded in the literature, some in equipment and materials, which themselves are embedded in the literature, and others in the tacit knowledge of individuals engaged in the research. Using knowledge pieces in well-understood ways corresponds to a search process labeled as *exploitation*. Using existing knowledge pieces in new ways corresponds to an *explorative* search process, which is more likely to lead to major

breakthroughs but also comes with a substantial risk of no or low impact (March, 1991). From this perspective, novel research is more closely associated with exploration.

Drawing on a combinatorial perspective of the research process, *novelty* can be defined as the recombination of pre-existing knowledge components in an unprecedented fashion. This combinatorial view of novelty has been embraced by scholars in various disciplines (Arthur, 2009; Burt, 2004; Mednick, 1962; Schumpeter, 1939; Simonton, 2004; Weitzman, 1998). For example, Nelson and Winter (1982) state that “the creation of any sort of novelty in art, science or practical life – consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence.” Romer (1994) and Varian (2009) also argue that new combinations of existing components provide a potentially huge source of important new discoveries. The ability to make new combinations of existing knowledge pieces is one reason that “outsiders” from other disciplines arguably can provide exceptional insights when they move from one field to another, as physicist Leo Szilard did, when he switched from physics to biology in the 1950s (Carroll, 2013, p. 352).

The combinatorial view of novelty has been studied in the technological invention literature and operationalized using patent information. Fleming (2001) takes the technology subclasses in which patents are classified as representing the components of technological know-how and defines inventors’ familiarity of a particular combination of subclasses as its occurrence in history weighted by time. Viewing more familiar combinations as less novel, he finds that novel combinations lead to lower average patent citations but a higher variance of citations.

Verhoeven, Bakker, and Veugelers (2016) combine this combinatorial novelty measure with a measure of novelty in technological and scientific knowledge origins, based on whether the focal patent cites other technological inventions or scientific literature from areas that were never cited before in its patent class. They find that the combination of the combinatorial novelty and the novelty in knowledge origins is a powerful identifier of breakthrough inventions.

Uzzi, Mukherjee, Stringer, and Jones (2013) apply a conceptually similar approach to scientific publications. They propose to trace the combinatorial process underlying the research from the references of the published paper. Operationally, they view journals as bodies of knowledge pieces and calculate the relative commonness for each pair of journals referenced by a paper. For this individual paper, they then use the lowest 10th percentile commonness score of its series of

commonness scores as an indication of its “novelty” and the median commonness score as an indication of its “conventionality.” They find that papers with both high novelty and conventionality are more likely to become top cited. Lee, Walsh, and Wang (2015) adapt the Uzzi et al. (2013) measure for their study of creativity in scientific teams and find that the effect of team characteristics on novelty is different from its effect on impact of the publication produced by the team.

Other approaches to assess combinatorial novelty in science also exist in the literature. In a field experiment conducted at a top American medical school, Boudreau, Guinan, Lakhani, and Riedl (2016) identify whether a research proposal departs from the existing literature, by examining all possible pairs of MeSH (Medical Subject Headings) terms in the proposal and then calculating the fraction of the pairs which have not appeared in all the previous literature in PubMed. They find that evaluators systematically give lower scores to highly novel research proposals.

Azoulay, Güler, Koçak, Murciano-Goroff, and Anttila-Hughes (2012) measure the recombinative character of a publication in a similar manner, examining the extent to which pairs of its MeSH descriptors are unusual. They find a negative association between the degree of recombinaiveness of a paper and the citation volume.

Taking a network perspective on science, novelty can be understood as making new connections or bridging structural holes in the network of science (Chen et al., 2009; Rzhetsky, Foster, Foster, & Evans, 2015; Shi, Foster, & Evans, 2015). Building on this network view of science, Klavans and Boyack (2013) cluster publications using co-citation analysis and then classify publications into four categories: *uniform*, *conform*, *innovate*, and *deviate*, based on the average distance between the clusters of referenced publications, as well as the focal publication. They observe that more innovative publications receive more citations. Foster, Rzhetsky, and Evans (2015) categorize five research strategies for biochemistry research: *jump* (introducing new chemicals), *new consolidation* (introducing new connections between chemicals in the same cluster), *new bridge* (introducing new chemical connections across clusters), *repeat consolidation* (repeating existing chemical connections within the same cluster), and *repeat bridge* (repeating existing chemical connections across clusters). Classifying the first three strategies as innovative ones, they find that, compared with conservative publications, innovative ones on average receive more

citations, have a higher standard deviation in citations, and are more likely to be among the top 1% highly cited publications and win biomedical or chemistry awards.

Following the combinatorial novelty approach, this paper assesses the novelty of a research article by examining the extent to which it makes novel combinations of prior knowledge components. In operationalizing the combinatorial novelty approach, we follow Uzzi et al. (2013) and use journals as bodies of knowledge components. Rather than looking at the *atypicality* of referenced journal pairs as do Uzzi et al. (2013), we focus specifically on the *novelty* of referenced journal pairs by examining whether a pair has never been made in prior publications and is thus *new*. Furthermore, we take into account the knowledge distance between the newly-combined journals based on their co-cited journal profiles, i.e., their common “friends”, to assess the difficulty of making the new combination. More precisely, we measure the novelty of a paper as the number of new journal pairs in its references weighted by the cosine similarity between the newly-paired journals.

It is important to note that combinatorial novelty is not the only way in which breakthroughs are made. For example, breakthroughs can result from a new observation coming to light, a completely new instrument becoming available, or the discovery of a new specie. It is also important to note that novelty (an ex ante character) is not identical to breakthrough (ex post, depending on success, usage, or impact). Not all breakthroughs result from novel research; many breakthroughs result from a series of cumulative and incremental research following on a novel idea. However, there is strong anecdotal evidence that research of a novel nature not only has the potential to become a breakthrough itself but also contributes to subsequent breakthroughs. The diagrams that Feynman produced in the late 1940s provided physicists with an entirely new way of understanding the behavior of subatomic particles and, according to the historian of physics David Kaiser, “revolutionized nearly every aspect of theoretical physics” (Kaiser, 2009, p. 4). The creation of transgenic and knockout mice in the late 1980s revolutionized research on any number of diseases. Or, consider the research of Sebastian Seung that has received considerable attention and aims at mapping the human brain, something that no one to date has done (Cook, 2015). Seung’s course is heavily influenced by applying a method described in a highly-cited paper published in *PLoS BIOLOGY* that used a novel approach in human connectome (Denk & Horstmann, 2004).

3. Measuring novelty of scientific publications

3.1.Procedure

We construct our novelty indicator for research articles published in 2001 and indexed in the Web of Science Core Collection (WoS), based on their references.

- For each paper, we retrieve all of its referenced journals and pair them up (i.e., J_1 - J_2 , J_1 - J_3 , J_1 - J_4 ...).
- We examine each journal pair to see whether it is new, i.e., has never appeared in prior literature starting from 1980¹.
- For those new journal pairs (e.g., J_1 - J_2), we assess how easy it is to make this new combination, by investigating how many common “friends” the paired journals have. More precisely, we compare the co-citation profiles of the two journals (J_1 and J_2) in the preceding three years (i.e., 1998-2000).
 - We use the following matrix where each row or column provides the co-citation profile for a journal. The i,j -th element in this symmetric matrix is the number of times that J_i and J_j are co-cited, that is, the number of papers published between 1998 and 2000 that cite the two journals together. For example, in the preceding three years, the pair J_1 and J_2 have never been cited together by any papers (as this pair is new), but J_1 and J_3 have been cited together by 3 papers, and J_2 and J_3 have been cited together by 6 papers, making J_3 a common friend of J_1 and J_2 , as is journal J_5 .

	J ₁	J ₂	J ₃	J ₄	J ₅	...
J ₁	/	0	3	0	5	...
J ₂	0	/	6	2	3	...
J ₃	3	6	/	5	4	...
J ₄	0	2	5	/	0	...
J ₅	5	3	4	0	/	...
...	/

¹ The 1980 cut off is because of data-availability reasons. It assumes a window of 20 years before obsolescence.

- The ease of combining J_1 and J_2 is then defined as the *cosine similarity* between their co-citation profiles:

$$COS_{1,2} = \frac{J_1 \cdot J_2}{\|J_1\| \|J_2\|}$$

where J_1 and J_2 are row (or column) vectors. *Cosine similarity* is a classic measure of similarity between two vectors and is widely used in bibliometrics.

- Correspondingly, the difficulty score of combining J_1 and J_2 is: $1 - COS_{1,2}$.
- For each paper, we construct a continuous measure of combinatorial *novelty* as the sum of the difficulty scores of making the new combinations. Papers without new combinations get 0 by definition.

$$Novelty = \sum_{J_i-J_j \text{ pair is new}} (1 - COS_{i,j})$$

- We also construct two alternative measures for robustness tests (details in *Appendix III*): the maximum novelty score which focuses exclusively on the novelty score of the most distant new journal pair and the weighted share of new journal pairs in all pairs, which is essentially a means of normalizing our novelty measure for the number of all journal pairs.
- In addition, to avoid trivial combinations, we focus only on the most important journal combinations, i.e., we exclude 50 percent of the least cited journals (based on the number of citations in the preceding three years received by all their publications starting from 1980)². To further reduce the likelihood of picking up trivial combinations, we impose as a condition that the new combination must be reused at least once in the next three years. We check the robustness of the main results to these choices in *Appendix III*.

3.2. Illustration

A novel contribution in 2001 in the biomedical field is the discovery by Dr. Peter Klein and colleagues that valproic acid inhibits histone deacetylase. At the time of the discovery, Dr. Klein was a Howard Hughes Medical Institute Investigator. The discovery was published in the

² The threshold for citations is 226.

Journal of Biological Chemistry entitled “Histone Deacetylase Is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen” (Phiel et al., 2001).

Valproic acid (VPA) is a short-chained fatty acid widely used for treating epilepsy and bipolar disorder. It is also a potent teratogen. However, how VPA actually works in any of these settings was unknown. A rich volume of knowledge had been accumulating in the literature about VPA in connection to epilepsy, bipolar disorder, and teratogen. By way of example, research existed on the possible pathway (but not the direct target) through which VPA can prevent seizure, a pathway (through activating Wnt-dependent gene expression) and several direct targets of lithium (the mainstay of therapy for bipolar disorder), as well as structural requirements for the teratogenic activity of VPA.

Klein and colleagues discovered the direct target of VPA by making a new connection between these existing pieces of knowledge and another piece of existing knowledge, specifically, histone acetylation (HDAC) is a negative regulator of gene transcription in multiple settings. Making this new connection led to the hypothesis that VPA inhibits HDAC and in turn activates Wnt-dependent gene expression. To test this hypothesis, Klein and colleagues ran a series of experiments, comparing effects of VPA with effects of trichostatin A (a well-characterized inhibitor of HDAC), as well as comparing VPA with other chemicals to rule out alternative possibilities.

This discovery not only contributed to fundamental knowledge but also suggested new possible targets for treating bipolar disorders. In addition, by connecting the discovery that VPA inhibits HDAC with another piece of knowledge that HDAC inhibitors can prevent proliferation and induce differentiation of various types of cancer cells, the discovery also provided a new possible therapy for treating cancer. It has sparked numerous studies of VPA as an anti-cancer drug.

This 2001 paper cites 42 WoS-indexed journals. Of all possible journal pairs (861), 9 journal pairs are new, using the procedure described supra. The new combination between the knowledge that HDAC is a negative regulator of gene transcription and other pieces of knowledge about VPA is reflected in the new journal pair between *Gene Expression* and other journals such as the *Journal of Clinical Psychiatry* and *Neuropsychopharmacology*. The novelty score for this paper is 6.89, which places this paper in the top 1% of novel papers in its field in

2001 (i.e., Biochemistry & Molecular Biology). This paper thus illustrates how research with a character of combinatorial novelty referenced journal pairs that are new.

The *Journal of Biological Chemistry* (where the paper was published) had an Impact Factor of 7.258 in 2001, which ranked it in the upper quartile in its subject categories, Biochemistry & Molecular Biology (more precisely 29 out of 308). This paper is also among the top 1% highly cited papers in its subject category. Papers citing it include several articles published in *Nature*, *Science*, and *PNAS*, some of which are top cited papers themselves. *Appendix I* describes the calculation of the novelty score for this novel paper in more detail.

4. Data and descriptive statistics

To explore the properties of novelty and its relationship with impact, we use a dataset consisting of all research articles³ in WoS published in 2001 from all the 251 subject categories. There are 785,324 articles in total, and 661,910 of them have references to at least two WoS journals. Among these 661,910 articles, 267 have no subject category information and therefore are excluded, and 269,870 articles have more than one subject category (up to six subject categories) and are counted multiple times. The final 2001 dataset used has 661,643 unique publications and 1,038,238 observations. Our findings are robust when we (1) only analyze papers with a single subject category or (2) reassign papers with multiple subject categories and papers in the category of “Multidisciplinary Sciences” to the majority subject category of their references.

We expect our measure to identify only a small minority of papers as novel, since the majority of research is of an exploitative rather than an exploratory nature. Indeed we find that relatively few papers make new referenced journal combinations. To be more specific, 89% of all papers in our sample make no new combinations of referenced journal and therefore do not score on the novelty measure. Of the 11% that make new journal combinations, most (54%) make only one new combination, and only 7% have more than 5 new combinations. Most of the novel papers score only modestly on our distance-weighted novelty indicator. At the other end of the distance

³ Since we are interested in original research, we keep only publications labeled as “article” in WoS but exclude other document types such as “review” and “letter.”

distribution, we find the top 10% most novel papers (within the set of novel papers) to have a score on our distance weighted indicator in the range of the interval (3.84-200.96).

Because our measure of novelty displays a highly skewed phenomenon of novelty in scientific publications, we construct a categorical novelty variable *NOV CAT*: (1) *non-novel*, if a paper has no new journal combinations, (2) *moderately novel*, if a paper makes at least one new combination but has a novelty score lower than the top 1% of its subject category, and (3) *highly novel*, if a paper has a novelty score among the top 1% of its subject category. We are particularly interested in papers which are highly novel.

Highly novel papers not only make more but also more distant new combinations. The median number of new combinations they make is 7, while the median for moderately novel papers is 1. The fact that the new combinations that highly novel papers make are more distant is suggested by their *cosine similarity* scores being lower than the scores of moderately novel papers (*Table 1*).

Insert Table 1 here

It is important to note the difference between novelty and interdisciplinarity (Larivière, Haustein, & Börner, 2015; Wang, Thijs, & Glänzel, 2015; Yegros-Yegros, Rafols, & D'Este, 2015). Not unexpectedly, new combinations are more likely to cross disciplinary boundaries: about 96% of the new journal combinations identified in our sample are cross-disciplinary, i.e., the newly paired journals do not share any common WoS subject categories. Nevertheless, crossing disciplines does not guarantee novelty: less than 8% of the cross-disciplinary journal combinations are new. In other words, while crossing disciplines is a source of novelty, most cross-disciplinary combinations are not novel. The novelty that we identify is a rarer activity in science than interdisciplinary research.

In addition, fields differ in their propensity to make new combinations. The Life Sciences score relatively higher on our novelty indicator, especially Neurosciences, Pharmacology and Biology & Biochemistry. The Physical Sciences score relatively lower on novelty, especially Space

Sciences and Physics. Social Sciences, especially Psychology, score above most fields⁴. Field difference in the novelty intensity may be partly explained by their heterogeneous patterns of publishing and referencing. Another possible explanation pertains to how research is conducted in the field; in some fields the research process may involve more combinative aspects than others. In the econometric analysis we control for scientific field (i.e., WoS subject category) specific effects.

5. Novelty and impact

5.1. High risk of novel research

In view of the risky nature of novel research, we expect novel papers to have a higher variance in their citation performance. Following Fleming (2001), the Generalized Negative Binomial (GNB) model is used to estimate the effects of novelty on the distribution characteristics of received citations. Specifically, GNB assumes that the number of citations (i.e., the dependent variable) follows a negative binomial distribution and allows us to model the natural logarithm of the mean μ and the natural logarithm of the dispersion parameter α each by a linear equation of novelty and other control variables. The variance of the distribution is $\sigma^2 = \mu + \alpha\mu^2$. For fitting the model the STATA function *gnbreg* is implemented (StataCorp, 2016).

We use a 15-year time window to count citations for our set of 2001 papers, which is deemed sufficiently long across fields (Wang, 2013). We control for other confounding factors with potential influence on the relationship between novelty and impact. First, we control for specific scientific field effects, by including the complete set of dummies for the 251 WoS subject categories. Second, we control for the number of references made in the focal paper, which might increase both the likelihood of having new combinations and the number of received citations (Bornmann, Leydesdorff, & Wang, 2014; Lee et al., 2015). Third, we take into account the size and nature of the collaborative effort, which might affect both novelty and impact (Adams, Black, Clemmons, & Stephan, 2005; Katz & Hicks, 1997; Lee et al., 2015).

⁴ By construction, there is no field differences in the relative share of highly novel (i.e., *NOV CAT* = 3) papers: *NOV CAT3* is defined as the top 1% novel papers within given subject categories.

Specifically, we include the number of authors and whether the paper is internationally coauthored as additional controls.

GNB model estimates are reported in *Table 2* and illustrated in *Figure 1A*. Of particular interest is the variance of the citation distribution. Results show that indeed highly novel papers have a much higher dispersion in citations; the dispersion of the citation distribution is 18% ($e^{0.162}-1$) higher for highly novel papers than non-novel papers. Moderately novel papers, however, do not differ significantly from non-novel ones, in terms of citation dispersion.

Insert Table 2 here

Insert Figure 1 here

A higher dispersion in impact can be driven by more extreme successes and/or more cases of uncited or rarely cited papers. Therefore, we examine in which tail of high and low impact the highly novel papers are more likely to be. We do this using multinomial logistic regression (*Table 3*). We classify papers within the same WoS subject category and publication year into three citation classes based on their citations in the 15-year time window: the top 10%, the lowest 10%, and the middle 80%. There is clear evidence that highly novel papers, which have a higher dispersion in their citations, are more likely to be in the tail of high impact. Specifically, the odds of being top 10% cited versus being middle 80% cited are 18% ($e^{0.162}-1$) higher for highly novel than non-novel papers. There is also strong evidence that highly novel papers are more likely to be in the tail of least cited papers: the odds of being in the lowest 10% cited versus being in the middle 80% cited are 15% ($e^{0.137}-1$) higher for highly novel than non-novel papers. In other words, the higher dispersion in citations for highly novel papers is driven by both tails of high and low impact and therefore reflects their higher level of uncertainty. On the other hand, moderately novel papers are only more likely to be in the top tail, not in the lower tail, displaying a lower level of uncertainty compared with highly novel papers, in line with the GNB results.

Insert Table 3 here

5.2.High gain from novel research

While novel research faces a higher level of risk, we also expect novel research to have a higher probability of making a significant contribution to research. We first examine whether novel papers are more likely to become “big hits,” i.e., receive an exceptionally large number of citations, defined here, following the bibliometric convention, as being top 1% highly cited in the same WoS subject category and publication year. We use the same 15-year citation time window to count citations as in previous analyses. Logistic regression controlling for previously mentioned other potential confounding factors reveals that the odds of a big hit are 57% ($e^{0.451}-1$) higher for highly novel papers and 13% ($e^{0.122}-1$) higher for moderately novel papers, compared with comparable non-novel papers (*Table 4* column 1 and *Figure 1B*).

Insert Table 4 here

Second, we find that novel papers are more likely to be cited by other big hits. Novel research is therefore not only more likely to become a big hit itself but also more likely to stimulate follow-on research which generates major impact. Specifically, we find that papers that cite novel papers are more likely to themselves receive more citations, compared with papers citing non-novel papers (*Appendix II Table A3*). Likewise, the probability of being cited by an article which itself becomes a big hit is higher for highly novel papers than for non-novel papers. We use a logistic model to estimate the probability of a paper being cited by big hits, teasing out any contamination from direct citations received, in addition to controlling for previously mentioned other confounding factors. We observe that the odds of being cited by big hits are 26% ($e^{0.229}-1$) higher for a highly novel paper than for comparable non-novel papers receiving the same number of citations (*Table 4* column 2 and *Figure 1C*)⁵. Compared with highly novel papers, moderately novel papers demonstrate a much smaller advantage over non-novel papers. The odds of being

⁵ In this analysis, big hits, which cite the focal paper, are identified as the top 1% highly cited papers in the same subject category and publication year, based on their cumulative citations till the end of 2015. Given that we do not have a sufficiently long time window to count citations for very recent papers, we only account for big hits between 2001 and 2010 and accordingly test whether novel papers are more likely to be cited by big hits in the 10-year period from 2001 to 2010. Correspondingly, we control for the number of direct citations in the same 10-year period.

cited by big hits are 6% ($e^{0.055}-1$) higher for a moderately novel paper than a comparable non-novel paper.

5.3. Transdisciplinary impact of novel research

We explore the disciplinary breadth of impact, that is, whether novel research is cited across more and more distant scientific fields than is non-novel research. We use the number of subject categories citing the focal paper in the 15 years after publication as the dependent variable and estimate a Poisson model, where we additionally control for the number of citations, given that papers with more citations are more likely to be cited by more fields. Results show that, compared with non-novel papers receiving the same number of citations and having the same values on all other control variables, highly and moderately novel papers are cited by 19% ($e^{0.177}-1$) and 11% ($e^{0.100}-1$) more subject categories, respectively (*Table 4* column 3 and *Figure 1D*).

We further examine whether the impact of novel papers reaches fields that are further away from their home field, compared with that of non-novel papers. First, we test whether the impact of a novel paper is more likely to be outside its home field than within its home field. To this end, we partition a paper's forward citations into two types: "home" and "foreign" field citations, that is, citations received from subsequent publications that share at least one common WoS subject category with the focal publication (home field citations) and citations from publication that share no common WoS subject categories (foreign field citations). Then we calculate, for each paper, the proportion of its citations that are foreign field citations. An OLS model (*Table 4* column 4) shows that novel papers have a larger share of citations from foreign fields. For papers which do have impact in foreign fields, we further investigate the distance between the citing foreign field and their home field. Specifically, we calculate, for each paper, the maximum distance between its home field and the foreign fields where it is cited. The pairwise distance between two fields is defined as $1 - \text{cosine similarity}$ between their co-citation profiles in the preceding three years. We find that this maximum distance between citing foreign field and home field is higher for novel than non-novel papers (*Table 4* column 5), suggesting that novel research has a greater

transdisciplinary impact reaching into more distant scientific domains than does non-novel research⁶.

The greater transdisciplinary impact of novel research raises the question of whether the major impact that novel papers generate is driven by their impact within and/or outside their home field. To answer this question we examine separately whether novel papers are among the top 1% highly cited by their home field and by foreign fields (*Table 4* columns 6-7 and *Figure 1E*)⁷. We find that the odds of being top cited in home fields are not significantly larger for highly novel papers than non-novel papers, and for moderately novel papers they are 11% ($e^{0.102}-1$) lower compared with that of non-novel papers. At the same time, novel papers, compared with non-novel ones, have much higher odds of being highly cited by foreign fields. Although this holds for moderately novel papers, it especially holds for highly novel papers, i.e., the odds of being highly cited in foreign fields are 37% ($e^{0.318}-1$) and 95% ($e^{0.669}-1$) higher for moderately and highly novel papers respectively, compared with that of non-novel papers. The finding that the overall high impact of novel research is due to its success in foreign fields rather than in its home field is consistent with resistance in the home field from existing paradigms against novel approaches and calls to mind the passage from Luke 4:24: “Verily I say unto you, No prophet is accepted in his own country.”

⁶ The findings that novel papers have an impact which is broader and more transdisciplinary (i.e., are cited by more fields and have a larger ratio of foreign field citations) are robust when we additionally control for the number of WoS subject categories that the focal paper itself is affiliated with.

⁷ It is important to note that a paper being highly cited in foreign fields means that, compared with other papers in the same home WoS subject category and publication year, its number of citations from foreign fields is among the top 1% of all citations from foreign fields to the home field. It does not mean that this paper is among the top 1% highly cited in a specific foreign field looking at all citations in the foreign field. To address this latter question, we have to use a different strategy. We first count each paper's citations from each of the 68 *subfields* (Glänzel & Schubert, 2003) separately and then identify the top 1% cited papers for each *subfield*, within the whole set of 2001 papers across all fields based on their citations received from this particular *subfield*. Subsequently we check whether a paper is among the top 1% cited in at least one of its foreign *subfields*. Logit regression, using the same setup as in *Table 4* column 7, shows that highly novel papers are significantly more likely to be a top cited paper in at least one foreign *subfield* compared with non-novel papers in that field. For moderately novel papers, no significant effects are found.

5.4. Delayed recognition for novel research

The major impact of novel research may take longer to realize because of resistance from existing paradigms or simply because it takes more time to incorporate novel research into subsequent research. To explore the extent to which delays in recognition occur, we estimate the probabilities of being a top 1% highly cited paper for non-, moderately-, and highly- novel papers for citation windows ranging from 1 to 15 years. We find that highly novel papers are less likely to be top cited when using citation time windows shorter than 3 years (*Table 5, Figure 1F, and Figure 2A*). As of the fourth year after publication, highly novel papers are significantly more likely to be top cited, and their advantage over non-novel papers increases with the length of the time window. Moderately novel papers suffer even more from delayed recognition. They are less likely to be top cited when using citation windows shorter than 5 years, and they only have a significantly higher chance of being a big hit with windows of at least 9 years.

Insert Table 5 here

Insert Figure 2 Here

The well-known fact that it takes longer for papers in one field to be cited in another field (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2002) raises the question of whether the finding of delayed recognition for novel research is driven by their large share of citations that come from foreign fields. We unravel the delayed recognition results further by comparing the time profile in recognition separately for home and foreign fields. A number of interesting results emerge.

First, the lower impact which novel papers face in their home field compared with non-novel papers shrinks over time, showing that delayed recognition for novel papers exists in their home field (*Appendix II Table A4 and Figure 2B*). More specifically, we find that highly novel papers, compared with non-novel papers, are significantly less likely to be top cited in their home field in the first seven years, but this disadvantage disappears when using a longer window. Moderately novel papers, however, are consistently, over time, significantly less likely to be top cited in their home field compared with non-novel papers. But also in this case, the gap with non-novel papers

in the probability of being top cited in the home field shrinks over time for moderately novel papers, just as it does for highly-novel papers.

Second, the higher impact of novel papers in foreign fields compared with non-novel papers magnifies over time, suggesting the delayed recognition for novel papers also exists in foreign fields (*Appendix II Table A5* and *Figure 2C*). We find that both moderately and highly novel papers are more likely to be highly cited in foreign fields than non-novel papers, but this advantage requires a citation time window of at least three years for both highly and moderately novel papers. Moreover, the foreign advantage of novel papers clearly increases over time.

Third, as *Appendix II Figure A1* shows, impact in foreign fields takes longer to materialize than that in home fields. For all papers, regardless of novelty, the average number of annual citations in foreign fields, compared with that of home fields, is smaller in the first seven years but greater in later years. This implies that it takes time for larger success of novel papers in foreign fields to compensate for their lack of advantage in their home fields. This is illustrated by *Appendix II Figure A2* which shows that it takes time for the advantage that novel papers enjoy in foreign fields to cancel out any disadvantage they have in home fields.

In sum, the overall delayed recognition for novel papers is a composite effect consisting of a delayed recognition both in home as well as in foreign fields and a delayed process in knowledge diffusion to other fields.

5.5. Bias against novelty

The finding of delayed recognition for novel research bears direct implications for the use of bibliometric indicators in science policy. As novel papers suffer from delayed recognition and need a sufficiently long citation time window before reaching major impact, bibliometric indicators which use short citation time-windows are biased against novelty.

In this section, we explore further how novel research performs on other popular bibliometric indicators. Specifically, we examine the *Journal Impact Factor*, probably the most influential indicator used (or abused) for assessing the “quality” of journals and their articles. We investigate whether novel papers, with their “high risk/high gain” nature, are more or less likely to be published in high Impact Factor journals. We find that although novel papers are published

on average in journals with higher Impact Factors, compared with non-novel ones (*Appendix II Table A2*), the Poisson regression, controlling for other confounding factors such as field differences, reveals that the Journal Impact Factor of moderately- and highly-novel papers is significantly and substantially lower (approximately 10% ($1-e^{-0.103}$) and 17% ($1-e^{-0.182}$) respectively) than comparable non-novel papers (*Table 6* and *Figure 3*). This finding—that novel papers are published in journals with Impact Factors lower than their non-novel counterparts, *ceteris paribus*—suggests that novel papers encounter obstacles in being accepted by journals holding central positions in science. Moreover, the negative association between novelty and Journal Impact Factor is not due to novel papers being more likely to be published in new journals. Regression analyses which additionally control for journal age or whether the journal is new confirms that the journals in which novel papers are published have a lower Impact Factor compared with the journals in which non-novel papers are published (*Table 5*). The increased pressure journals are under to boost their Impact Factor (Martin, 2016) and the fact that the Journal Impact Factor is based on citations in the first two years after publication⁸ suggests that journals may strategically choose to not publish novel papers which are less likely to be highly cited in the short run.

Insert Table 6 here

Insert Figure 3 here

Another question is whether the negative association between novelty and Journal Impact Factor is responsible for the delayed recognition faced by novel research. To address this question, we examine whether the novelty effect on the probability of big hits is contingent on the Impact Factor of the journal in which the paper is published. If publication in a low Impact Factor journal is responsible for the delayed recognition encountered by novel papers, we expect that novel papers which succeed in getting in high Impact Factor journals would not suffer from delayed recognition. Therefore, we re-estimate the models in *Table 5*, additionally controlling for the Journal Impact Factor and incorporating interaction effects between novelty and whether

⁸ Journal Impact factor is essentially the average number of citations received in the current year by papers published in the preceding two years. <http://wokinfo.com/essays/impact-factor/>

the journal in which the focal paper is published has a top 10% Impact Factor in its subject category. As shown in *Appendix II Table A6* and *Figure 4*, novel papers published in high Impact Factor journals still have a delayed citation accumulation process compared with non-novel papers in high Impact Factor journals. We conclude that delayed recognition is not entirely due to publication of novel works in journals with lower than expected Impact Factors.

Insert Figure 4 here

5.6. Novelty and quality

Our research demonstrates that commonly used bibliometric indicators, specifically the Journal Impact Factor and others using short-term citation counts, are biased against novel papers. One might argue that such “bias” simply reflects the low quality associated with novel research. This raises the potential issue concerning unobserved and uncontrolled heterogeneity in paper quality. If novel research is associated with low quality, this would indeed explain the observation that novel papers are less likely to be highly cited in the short run and are less likely to be published in high Impact Factor journals, but it cannot explain why novel papers are more likely to eventually become highly cited and be cited in more fields. On the other hand, if novel research is associated with high quality, then it would explain its long-term big impact but not its delayed recognition, or its lower Journal Impact Factor, or the fact that novel papers which are published in high Impact Factor journals still display a delayed recognition. Although we cannot completely rule out the possible link between novelty and quality, due to the lack of a proper measure for the true quality of a paper, the citation patterns of novel research that we find in this paper suggest something different than a clear association between novelty and quality. Therefore, we can at least conclude that novelty affects *ex post* impact in a non-trivial fashion which is difficult to explain by its intrinsic quality.

5.7. Robustness analysis

We ran a set of robustness tests on our findings. Details are reported in *Appendix III*. First, we tested whether our findings are robust across scientific fields. All our findings are robust for hard sciences and engineering, but several findings are not robust for social sciences and humanities. Specifically, findings that novel research has a higher dispersion in citations, a lower probability

of being top 1% cited in the short run, and lower journal impact factors are not robust for arts and humanities, and the finding that novel research has a higher dispersion is not robust for social sciences. Although this may suggest that our findings hold only for hard sciences, it is more likely due to the insufficient coverage of WoS for humanities and social sciences (Hicks, 2004).

The dataset consists of 661,643 unique publications and 1,038,238 observations, where papers with multiple WoS subject categories are counted multiple times. We tested two alternative approaches: (1) excluding papers with multiple subject categories from the analysis or (2) reassigning papers with multiple subject categories and papers in the category of “Multidisciplinary Sciences” to the majority subject category of their references. All results are robust to both alternative approaches.

We also examined whether our findings are sensitive to variations of our novelty measure, which is essentially a distance-weighted number of new combinations. We tested two alternative formulations, i.e., (1) the maximum novelty score which focuses exclusively on the novelty score of the most distant new journal pair and (2) the weighted share of new journal pairs in all pairs, which is essentially a means of normalizing our novelty measure for the number of all journal pairs. Results are consistent when using these alternative formulas.

Our novelty measure excluded 50% of the least cited journals and required that the new combination of journals is reused in the next three years. Relaxing these constraints yielded robust results. Our results are also robust to additional constraints, such as excluding top 10% highly cited journals and multidisciplinary journals.

We used categorical novelty measures in our regression analysis (i.e., highly-, moderately-, and non-novel). We duplicated the results using the natural logarithm transformed continuous novelty score in the regression and obtained robust results. We classified papers with the highest 1% novelty score as highly novel. Using alternative thresholds, i.e., top 0.1% and 5%, also yielded consistent results.

Third, all our findings remain consistent and significant when we additionally control for the Uzzi et al. (2013) measure of *atypicality* in the regressions. More importantly, compared with the

atypicality measure, our novelty measure behaves more reliably and captures more the “high risk” nature of novel research.

6. Discussion

This research has a number of limitations. First, we focus on combinatorial novelty, which is only one possible approach for characterizing novelty. Novelty is an abstract and complex concept, easy to intuit but hard to define. Novelty has multiple dimensions or types, and our analysis only captures one of them. Second, we follow Uzzi et al. (2013) in viewing journals as bodies of knowledge and construct our novelty measure based on new combinations of journals in the references. Other strategies exist for identifying knowledge components, such as the keywords, or topics. Papers can also be clustered based on text-similarly, co-citations or bibliographic coupling.

This paper also raises a number of interesting research questions for future research. First, who is more likely to produce novel research: juniors or seniors, males or females, researchers at prestigious universities or those from peripheral institutions? Second, to what extent do funding agencies select novel proposals to support, and do certain funding models encourage funding recipients to take a more exploratory approach? Third, future research could examine the dynamic citation process and identify critical moments and mechanism triggering the diffusion of novel ideas. In addition to well documented *sleeping beauties*, there are many other general types of citation aging (Costas, Van Leeuwen, & Van Raan, 2010; Costas, van Leeuwen, & van Raan, 2013; Zhang, Wang, & Mei, 2017). Future research could investigate the kind of citation pattern that novel research typically follows. Fourth, it would be interesting to understand what kind of journals are more likely to accept novel research, and what are the mechanisms underlying this observed negative association? Is it because their editors strategically choose not to publish novel papers, anticipating their lower citation profile in the short run which would lower the Impact Factor of the journal, or because their peer review is conservative and tends to be biased against novelty (Boudreau et al., 2016; Horrobin, 1990), or because researchers choose not to submit their novel papers to high Impact Factor journals? Fifth, given that scientific disciplines are heterogeneous in their research processes and their social structure, which characteristics explain field differences in the propensity to generate combinatorial novelty, as well as how novelty is related to impact.

7. Conclusions

We propose that a way to measure the potential an article has to advance the knowledge frontier is to examine the combinatorial novelty of its references. To this end, we apply a newly minted measure of novelty to all WoS research articles published in 2001 across all scientific disciplines. We find that novel papers, in particular highly novel papers, exhibit citation patterns consistent with the “high risk/high gain” profile associated with breakthrough research. Novel papers have a significantly higher variance in citation performance than do non-novel papers, confirming their risky profile. At the same time, novel papers are associated with big hits. They have a significantly higher chance of being top 1% highly cited, and are more likely to lead to follow-up high impact research. Novel papers also have a broader impact across scientific fields, and are more likely to be highly cited in more distant foreign fields compared with non-novel papers. The big impact of novel papers comes from foreign fields, as novel papers are not more likely to be highly cited in their home fields. Furthermore, novel papers require a sufficient period of time before their important contribution is recognized. This delayed recognition is suggestive of reluctance from incumbent scientific paradigms to recognize novel approaches and the longer time period needed to incorporate novel research into subsequent research, particularly from other distant fields.

Delayed recognition leads novel research to perform poorly on bibliometric measures which use short citation windows. Novel research is published in journals with lower than expected Impact Factors, another widely (ab)used bibliometric measure. Moreover, even if novel research succeeds in being published in high impact factor journals, it still suffers from delayed recognition.

Taken together, our results suggest that some widely used bibliometric measures are biased against novel research and thus may fail to identify papers and individuals doing novel research. This bias against novelty imperils scientific progress, because novel research, as we have shown, is much more likely to become a big hit in the long run, particularly in fields other than their own, as well as to stimulate follow-up big hits.

The bias against novelty is of particular concern given the increased reliance funding agencies place on readily available bibliometric indicators in making funding and evaluation decisions.

The bias against novel papers may also help explain why funding agencies which increasingly rely on such measures are widely perceived as being more and more risk-averse, choosing “safe” projects over those that involve a higher level of uncertainty with regard to possible outcomes. In this respect, our research is consistent with that of Boudreau et al. (2016), who find that evaluators give lower scores to proposals that are highly novel where novelty is measured in terms of the proposal’s use of novel combinations of MeSH terms relative to the underlying literature.

The bias against novelty applies not only to funding decisions but to science policy more generally. The prevailing (mis)use of indicators which rely on short citation time windows and Journal Impact Factor in various decisions (e.g., hiring and tenure of researchers) at various levels (i.e., department, university, and national) is likely to disincentivize novel research. We advocate the awareness of such potential bias and suggest, when relying on bibliometric indicators, to use a wider portfolio of indicators and to adopt time windows beyond two or three years. Because novel research requires a long time window to reveal its full potential, assessing novelty for junior researchers is particularly problematic as their publications do not have a sufficiently long time window to accumulate citations.

In addition, the finding that novel papers, which typically cross disciplinary boundaries when venturing into novel approaches, are significantly more likely to become highly cited in foreign fields but not in their home field highlights the importance of avoiding a monodisciplinary approach in peer review. Peer review is widely implemented in science decision-making. It is typically organized along disciplinary lines, with peers within the same discipline making a judgment on the value of the research that is being evaluated. Studies of interdisciplinary research demonstrate that a discipline-based science system is detrimental to the advancement and societal accountability of science (Gibbons, 1994; The National Academies, 2004). This paper contributes to this discussion, suggesting that peer review which is bounded by disciplinary borders may fail to recognize the full value of novel research, which is typically cross-disciplinary in its origins and has its major impact realized outside its home field.

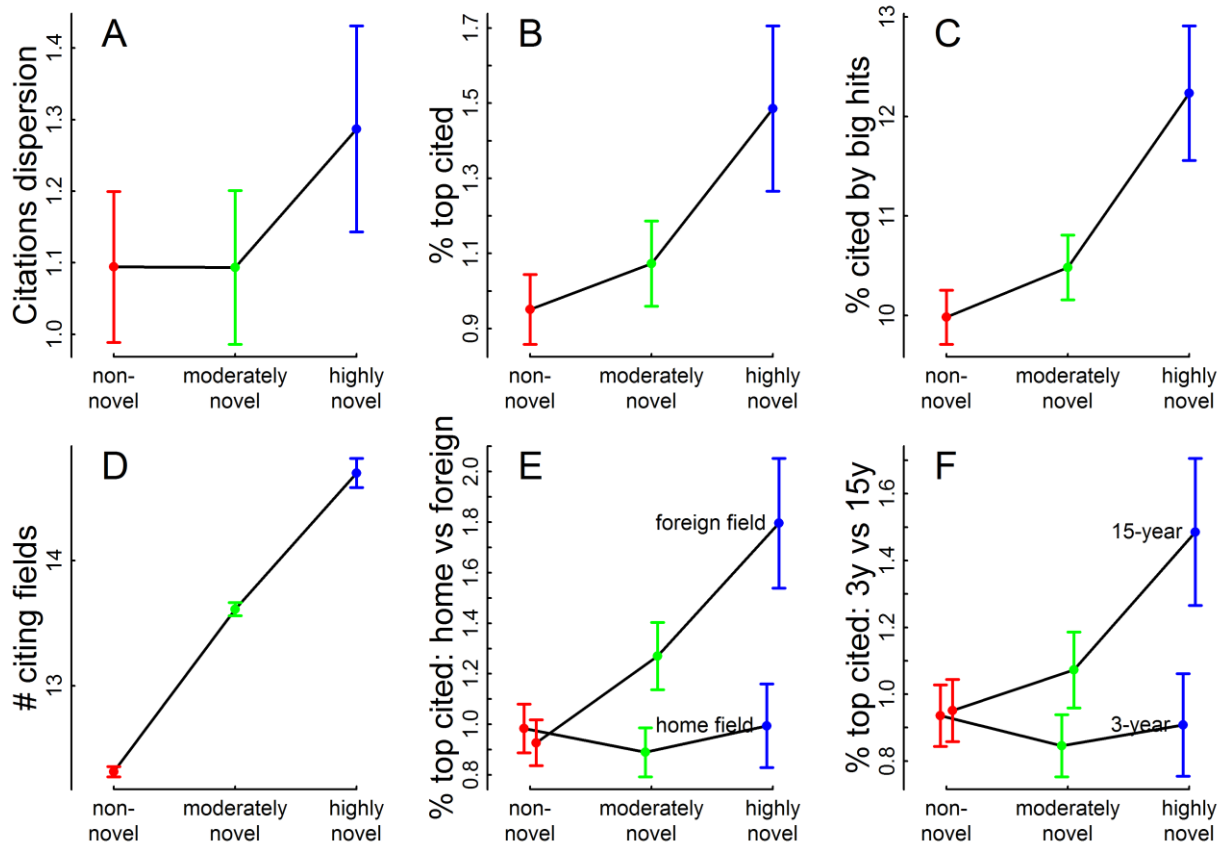


Figure 1. Impact profile of novel research. (A) Estimated dispersion of citations (15-year), based on the Generalized Negative Binomial model in *Table 2* column 3. (B) Estimated probability of being among the top 1% cited articles in the same WoS subject category and publication year, based on 15-year citations and the logit model in *Table 4* column 1. (C) Estimated probability of being cited by big hits, based on the logit model in *Table 4* column 2. (D) Estimated number of WoS subject categories citing the focal paper (15-year), based on the Poisson model in *Table 4* column 3. (E) Estimated probability of being among the top 1% cited articles in the same WoS subject category and publication year, based on 15-year home- and foreign-field citations separately. Estimations are based on two logit models in *Table 4* column 6 and 7. (F) Estimated probability of being among the top 1% cited articles in the same WoS subject category and publication year, based on 3-year⁹ and 15-year citations separately. Estimations are based on two logit models in *Table 5* column 3 and 15. All estimated values are for an average paper (i.e., in the biggest WoS subject category, not internationally coauthored, and with all other covariates at their means) in different novel classes. The vertical bars represent the 95% confidence interval. Data consist of 1,038,238 observations of 661,643 unique WoS articles published in 2001 and are sourced from Web of Science Core Collection.

⁹ For identifying the top 1% cited papers, we first extract the 99th percentile of the citation distribution for a field, and then classify a paper as top 1% cited if it has more than (not including) this number of citations. Therefore, there is normally less than 1% papers identified as big hits, in particular in the first few years after publication.

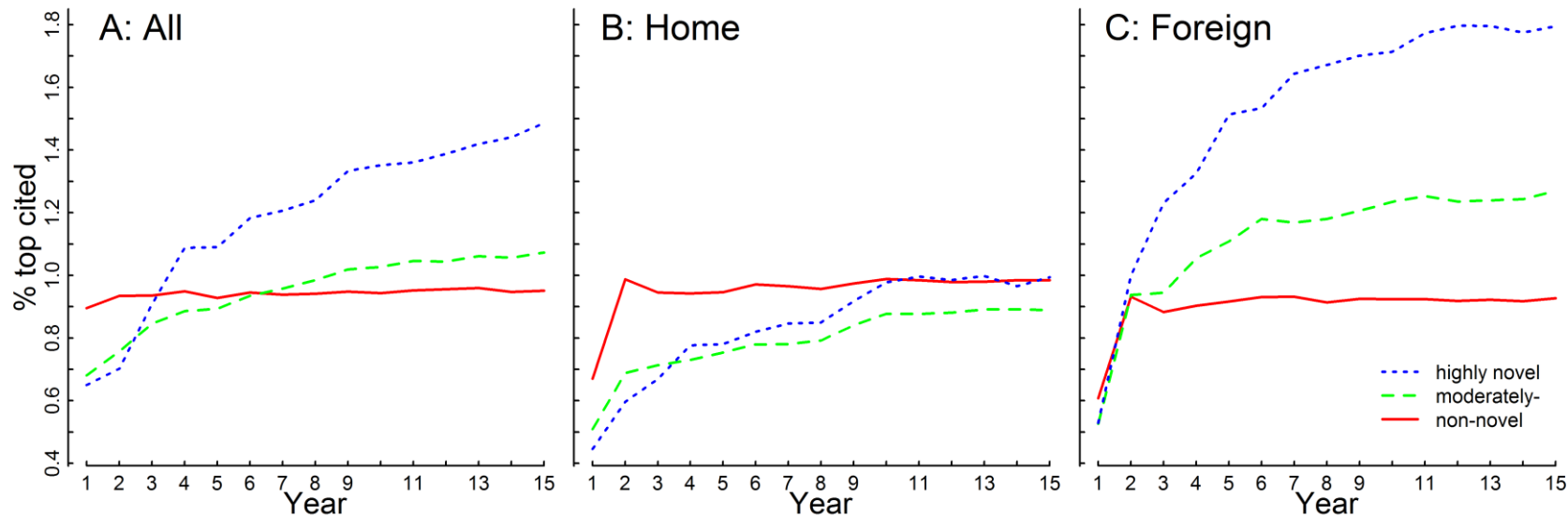


Figure 2. Citation dynamics and novelty. (A) Estimated probability of big hits, using 15 consecutive time windows to dynamically identify big hits. As an example, big hits in year 3 are identified as the top 1% highly cited papers based on their cumulative citations in a 3-year time window, i.e., from 2001 to 2003. Results are based on 15 logistic models reported in *Table 5*. (B) Estimated probability of big hits, based on home field citations, i.e., citations received from papers sharing at least one common WoS subject category with the focal cited paper. Results are based on 15 logistic models reported in *Table A4*. (C) Estimated probability of big hits, based on foreign field citations, i.e., citations received from papers sharing no common WoS subject categories with the focal cited paper. Results are based on 15 logistic models reported in *Table A5*. Data consist of 1,038,238 observations of 661,643 unique WoS articles published in 2001 and are sourced from Web of Science Core Collection.

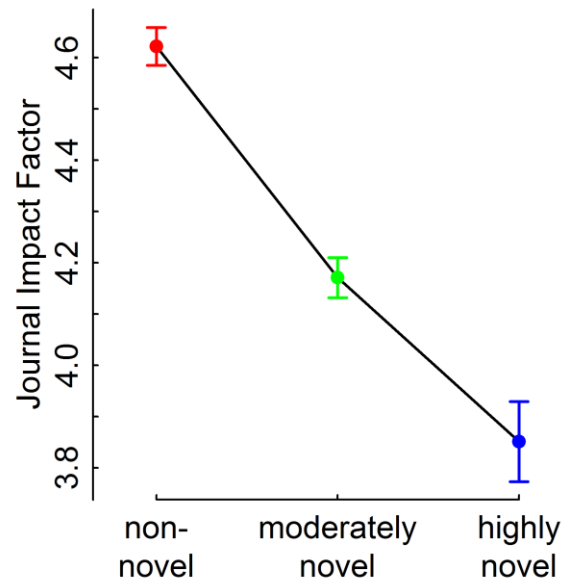


Figure 3. Journal Impact Factor and novelty. Estimated Journal Impact Factor for an average paper with different novelty classes, based on the Poisson model reported in *Table 6* column 3. Data consist of 1,038,238 observations of 661,643 unique WoS articles published in 2001 and are sourced from Web of Science Core Collection.

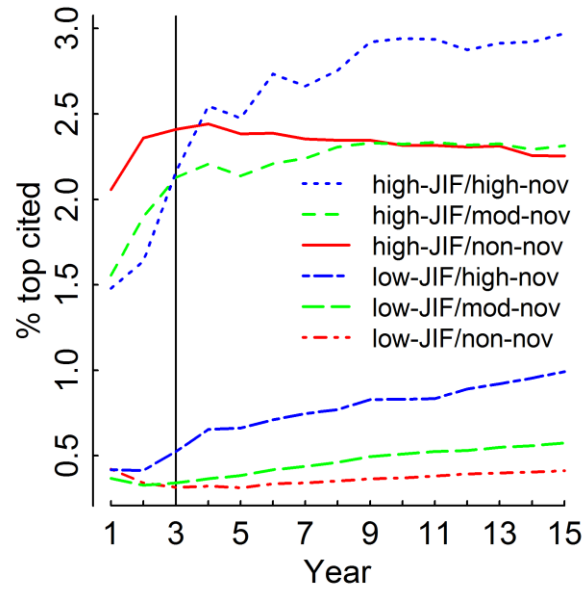


Figure 4. Citation dynamics and novelty, by JIF groups. Estimated probability of being a big hit by year, for papers in different novelty classes and Journal Impact Factor groups. Estimations are based on a set of logistic models additionally incorporating interaction effects between novelty classes and whether a journal has an Impact Factor among the top 10% in its field. Regression outputs are reported in *Appendix II Table A6*. Data consist of 1,038,238 observations of 661,643 unique WoS articles published in 2001 and are sourced from Web of Science Core Collection.

Table 1. Occurrence of novelty

	(1)	(2)	(3)	(4)	(5)	(6)
	# papers	% papers	Avg (avg cos)	Avg(min cos)	Avg # new pairs	Median # new pairs
Non-novel	919,333	88.55%	/	/	/	/
Moderately novel	108,635	10.46%	0.22	0.19	1.76	1.00
Highly novel	10,270	0.99%	0.13	0.06	8.39	7.00

Data sourced from Web of Science Core Collection.

Table 2. Mean and dispersion of citations

	Citations (15-year)		
	GNB		
	(1)	(2)	(3)
Mean			
ln(novelty+1)	0.051*** (0.006)		
Novel (dummy)		0.042*** (0.006)	
NOV CAT2			0.032*** (0.006)
NOV CAT3			0.146*** (0.019)
International	0.077*** (0.005)	0.077*** (0.005)	0.077*** (0.005)
ln(# authors)	0.264*** (0.005)	0.264*** (0.005)	0.264*** (0.005)
ln(# refs)	0.629*** (0.007)	0.631*** (0.006)	0.629*** (0.006)
Dispersion			
ln(novelty+1)	0.044*** (0.008)		
Novel (dummy)		0.015* (0.007)	
NOV CAT2			-0.001 (0.008)
NOV CAT3			0.162*** (0.023)
International	-0.060*** (0.007)	-0.061*** (0.007)	-0.060*** (0.007)
ln(# authors)	-0.144*** (0.006)	-0.144*** (0.006)	-0.144*** (0.006)
ln(# refs)	-0.244*** (0.008)	-0.239*** (0.008)	-0.242*** (0.008)
pubs.	661643	661643	661643
obs.	1038238	1038238	1038238
Pseudo R2	0.024	0.024	0.024
Log lik	-4333075	-4333181	-4333049

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** p<.001, ** p<.01, * p<.05, + p<.10.

Table 3. Citation classes

	Citation classes (15-year) Multinomial logit
Top 10% vs. middle 80%	
NOV CAT2	0.056*** (0.010)
NOV CAT3	0.162*** (0.026)
International	0.080*** (0.008)
ln(# authors)	0.444*** (0.006)
ln(# refs)	1.050*** (0.007)
Low 10% vs. middle 80%	
NOV CAT2	-0.054*** (0.015)
NOV CAT3	0.137* (0.056)
International	-0.225*** (0.012)
ln(# authors)	-0.585*** (0.007)
ln(# refs)	-1.077*** (0.007)
Pubs	661643
obs	1038238
Pseudo R2	0.084
Log lik	-540344

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. We classify papers, within the same WoS subject category and publication year, into three classes based on their 15-year citations: low 10%¹⁰, middle 80%, and top 10%. For the multinomial logit regression, the middle 80% is used as the reference group. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** p<.001, ** p<.01, * p<.05, + p<.10.

¹⁰ For identifying the least 10% cited papers, we first extracted the 10th percentile of the citation distribution for a field, and then classified papers with less than (not including) this number of citations as the least 10% cited. Therefore, if a field has more than 10% papers with 0 citations, no papers in this field are classified as the 10% least cited papers, and this field is automatically dropped from the regression analysis. In total 50 subject categories (68,481 unique publications and 121,977 observations) are dropped. As a robustness test, when a field has more than 10% papers with 0 citations, we randomly assign 10% as 10% least cited. The results are robust.

Table 4. Impact profile of novel research

	(1) Top 1% cited (15-year)	(2) Cited by big hits (10-year)	(3) # citing fields (15-year)	(4) Ratio foreign field citations (15-year)	(5) Max distance between citing foreign and home field (15-year)	(6) Top 1% cited home field (15-year)	(7) Top 1% cited foreign field (15-year)
	logit	logit	Poisson	OLS	OLS	logit	logit
NOV CAT2	0.122*** (0.029)	0.055*** (0.010)	0.100*** (0.001)	0.050*** (0.001)	0.016*** (0.000)	-0.102** (0.030)	0.318*** (0.028)
NOV CAT3	0.451*** (0.060)	0.229*** (0.029)	0.177*** (0.004)	0.083*** (0.003)	0.030*** (0.001)	0.010 (0.071)	0.669*** (0.056)
ln(10-year citations)		1.669*** (0.005)					
ln(15-year citations)			0.494*** (0.001)	0.002*** (0.000)			
ln(15-year foreign field citations)					0.052*** (0.000)		
International	0.078** (0.024)	0.005 (0.009)	-0.013*** (0.001)	-0.006*** (0.001)	-0.001* (0.000)	0.070** (0.024)	0.054* (0.025)
ln(# authors)	0.546*** (0.019)	-0.068*** (0.006)	-0.002* (0.001)	0.012*** (0.001)	-0.010*** (0.000)	0.505*** (0.019)	0.523*** (0.018)
ln(# refs)	1.198*** (0.021)	-0.101*** (0.007)	0.002* (0.001)	0.002*** (0.001)	-0.010*** (0.000)	1.218*** (0.021)	1.138*** (0.021)
pubs.	661414	610270	621595	621595	560290	661414	661414
obs.	1037695	962615	979817	979817	881201	1037695	1037695
(Pseudo) R2	0.048	0.306	0.465	0.134	0.360	0.044	0.047
Log lik	-54558	-283668	-2508306	/	/	-54503	-54324

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Top 1% cited (column 1) means being among the top 1% highly cited within the same WoS subject category and publication year, based on the number of citations in a 15-year time window between 2001 and 2015. Cited by big hit (column 2) means cited by an article, which is published in between 2001 and 2010 and among the top 1% highly cited articles within the same WoS subject category and publication year, based on their cumulative citations up to 2015. Big hits published after 2010 are not analyzed because their available time windows are too short to identify big hits reliably. # citing fields (column 3) is the number of WoS subject categories citing the focal paper. Ratio foreign field citations (column 4) is the proportion of all citations that are from papers sharing no common WoS subject categories with the focal cited paper. Max distance between the citing foreign and the home field (column 5) is the maximum distance between a paper's citing foreign field and its home field, where pairwise distance between two fields are calculated as $1 - \text{cosine similarity}$ between these two fields based on their co-citations in the preceding three year. Top 1% cited home field (column 6) means being among the top 1% highly cited within the same WoS subject category and publication year, based on 15-year citation received from papers that share at least one common WoS subject category with the focal cited paper. Top 1% cited foreign field (column 7) means being among the top 1% highly cited within the same WoS subject category and publication year, based on 15-year citation received from papers that share no common WoS subject category with the focal cited paper. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** p<.001, ** p<.01, * p<.05, + p<.10.

Table 5. Big hits and novelty: Delayed recognition

	Top 1% cited (1y) logit	Top 1% cited (2y) Logit	Top 1% cited (3y) logit	Top 1% cited (4y) logit	Top 1% cited (5y) logit	Top 1% cited (6y) logit	Top 1% cited (7y) logit	Top 1% cited (8y) logit	Top 1% cited (9y) logit	Top 1% cited (10y) Logit	Top 1% cited (11y) logit	Top 1% cited (12y) logit	Top 1% cited (13y) logit	Top 1% cited (14y) logit	Top 1% cited (15y) logit
NOV CAT2	-0.28*** (0.04)	-0.21*** (0.03)	-0.10** (0.03)	-0.07* (0.03)	-0.04 (0.03)	-0.01 (0.03)	0.02 (0.03)	0.05 (0.03)	0.07* (0.03)	0.09** (0.03)	0.10** (0.03)	0.09** (0.03)	0.10*** (0.03)	0.11*** (0.03)	0.12*** (0.03)
NOV CAT3	-0.32** (0.10)	-0.29** (0.08)	-0.03 (0.07)	0.14* (0.07)	0.16* (0.07)	0.23*** (0.06)	0.25*** (0.06)	0.28*** (0.06)	0.34*** (0.06)	0.36*** (0.06)	0.36*** (0.06)	0.38*** (0.06)	0.40*** (0.06)	0.42*** (0.06)	0.45*** (0.06)
International	0.22*** (0.03)	0.19*** (0.02)	0.14*** (0.02)	0.13*** (0.02)	0.11*** (0.02)	0.11*** (0.02)	0.11*** (0.02)	0.09*** (0.02)	0.10*** (0.02)	0.10*** (0.02)	0.08** (0.02)	0.07** (0.02)	0.06* (0.02)	0.08** (0.02)	0.08** (0.02)
ln(# authors)	0.61*** (0.02)	0.76*** (0.02)	0.83*** (0.02)	0.80*** (0.02)	0.78*** (0.02)	0.75*** (0.02)	0.73*** (0.02)	0.70*** (0.02)	0.67*** (0.02)	0.65*** (0.02)	0.62*** (0.02)	0.60*** (0.02)	0.58*** (0.02)	0.56*** (0.02)	0.55*** (0.02)
ln(# refs)	1.02*** (0.02)	1.25*** (0.02)	1.29*** (0.02)	1.28*** (0.02)	1.29*** (0.02)	1.28*** (0.02)	1.27*** (0.02)	1.25*** (0.02)	1.23*** (0.02)	1.23*** (0.02)	1.23*** (0.02)	1.22*** (0.02)	1.21*** (0.02)	1.20*** (0.02)	1.20*** (0.02)
pubs.	660282	661387	661414	661414	661414	661271	661414	661414	661414	661414	661414	661413	661414	661414	661414
obs.	1033393	1037232	1037695	1037695	1037228	1037465	1037695	1037695	1037695	1037695	1037695	1037541	1037695	1037695	1037695
Pseudo R2	0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Log lik	-40027	-48961	-50917	-52168	-52708	-53156	-53208	-53790	-53899	-54023	-54250	-54322	-54411	-54429	-54558

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Big hits using a t -year time window means among the top 1% highly cited papers within the same WoS subject category and publication year, based on their citations in the t -year time window starting from 2001.

Identifying big hits using annual citation counts instead of cumulative citation counts yields similar results. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$.

Table 6. Journal Impact Factor and novelty

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Journal age	Jnl Age <4	JIF	JIF	JIF	JIF TOP 10%	JIF TOP 10%	JIF TOP 10%
	Poisson	logit	Poisson	Poisson	Poisson	logit	logit	logit
NOV CAT2	-0.078*** (0.003)	0.130*** (0.015)	-0.103*** (0.003)	-0.079*** (0.003)	-0.101*** (0.003)	-0.196*** (0.009)	-0.144*** (0.009)	-0.193*** (0.009)
NOV CAT3	-0.171*** (0.008)	0.367*** (0.041)	-0.182*** (0.010)	-0.136*** (0.009)	-0.180*** (0.010)	-0.406*** (0.028)	-0.293*** (0.028)	-0.403*** (0.028)
ln(journal age)				0.250*** (0.001)			0.791*** (0.005)	
Journal age < 4					-0.398*** (0.007)			-0.677*** (0.025)
International	0.034*** (0.002)	-0.082*** (0.013)	0.071*** (0.002)	0.062*** (0.002)	0.070*** (0.002)	0.125*** (0.007)	0.107*** (0.007)	0.124*** (0.007)
ln(# authors)	0.042*** (0.001)	-0.164*** (0.008)	0.171*** (0.002)	0.158*** (0.002)	0.170*** (0.002)	0.443*** (0.005)	0.428*** (0.005)	0.441*** (0.005)
ln(# refs)	0.106*** (0.001)	-0.106*** (0.008)	0.347*** (0.002)	0.319*** (0.002)	0.346*** (0.002)	0.879*** (0.005)	0.835*** (0.006)	0.878*** (0.005)
pubs.	661643	653339	642411	642411	642411	639334	639334	639334
obs.	1038238	1012721	1006687	1006687	1006687	994369	994369	994369
Pseudo R2	0.110	0.051	0.196	0.212	0.197	0.105	0.142	0.106
Log lik	-8001948	-199522	-1659741	-1625885	-1656957	-415009	-397579	-414548

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Journal age is the year of publication (i.e., 2001 of our sample) minus the first year that a journal started publishing and then plus 1. In other words, a journal has an age of 1 in its first year of publishing. Results of Poisson models are reported here, an alternative specification, using the OLS model and the log of journal age or JIF as the dependent variable yields consistent results. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** p<.001, ** p<.01, * p<.05, + p<.10.

References

- Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from US universities, 1981-1999. *Research Policy*, 34(3), 259-285. doi:10.1016/j.respol.2005.01.014
- Alberts, B. (2010). Overbuilding Research Capacity. *Science*, 329(5997), 1257-1257. doi:10.1126/science.1197077
- Arrow, K. (1962). Economic welfare and the allocation of resources for invention *The rate and direction of inventive activity: Economic and social factors* (pp. 609-626). Princeton, NJ: Princeton University Press.
- Arthur, W. B. (2009). *The nature of technology : what it is and how it evolves*. New York, NY: Free Press.
- Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3), 527-554. doi:10.1111/j.1756-2171.2011.00140.x
- Azoulay, P., Güler, I., Koçak, Ö., Murciano-Goroff, R., & Anttila-Hughes, J. (2012). *Are recombinant scientific articles more impactful?*
- Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, 8(1), 175-180. doi:10.1016/j.joi.2013.11.005
- Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. *Management Science*. doi:doi:10.1287/mnsc.2015.2285
- Boyack, K. W., & Klavans, R. (2014). Atypical combinations are confounded by disciplinary effects. *STI 2014 Leiden*, 64.
- Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 349-399. doi:10.1086/421787
- Butler, L. (2003). Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Research Policy*, 32(1), 143-155. doi:10.1016/s0048-7333(02)00007-0
- Carroll, S. B. (2013). *Brave genius : a scientist, a philosopher, and their daring adventures from the French resistance to the Nobel Prize* (First edition. ed.). New York, NY: Crown Publishers.
- Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191-209. doi:<http://dx.doi.org/10.1016/j.joi.2009.03.004>
- Cook, G. (2015, January 11). Sebastian Seung's Quest to Map the Human Brain. *The New York Times Magazine*, MM27.
- Costas, R., Van Leeuwen, T., & Van Raan, A. (2010). Is scientific literature subject to a 'Sell By Date'? A general methodology to analyze the 'durability' of scientific documents. *Journal of the American Society for Information Science and Technology*, 61(2), 329-339.
- Costas, R., van Leeuwen, T. N., & van Raan, A. F. J. (2013). Effects of the durability of scientific literature at the group level: Case study of chemistry research groups in the Netherlands. *Research Policy*, 42(4), 886-894. doi:<http://dx.doi.org/10.1016/j.respol.2012.11.006>
- Denk, W., & Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *Plos Biology*, 2(11), 1900-1909. doi:10.1371/journal.pbio.0020329
- Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1), 117-132. doi:10.1287/mnsc.47.1.117.10671
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review*, 80(5), 875-908. doi:doi:10.1177/0003122415601618

- Garfield, E. (1980). Premature discovery or delayed recognition—Why. *Current Contents*, 21, 5-10.
- Gibbons, M. (1994). *The new production of knowledge*. London, UK: Sage Publications.
- Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357-367. doi:10.1023/A:1022378804087
- Hicks, D. (2004). The Four Literatures of Social Science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473-496). New York, NY: Springer.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251-261. doi:10.1016/j.respol.2011.09.007
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429-431.
- Horrobin, D. F. (1990). The philosophical basis of peer review and the suppression of innovation. *JAMA*, 263(10), 1438-1441. doi:10.1001/jama.1990.03440100162024
- Kaiser, D. (2009). *Drawing theories apart: The dispersion of Feynman diagrams in postwar physics*. Chicago, IL: University of Chicago Press.
- Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541-554. Retrieved from <Go to ISI>://000071723600016
- Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, 125(5), 524.
- Klavans, R., & Boyack, K. (2013). *Towards the development of an article-level indicator of conformity, innovation and deviation*. Paper presented at the 18th International Conference on Science and Technology Indicators.
- Kolata, G. (2009, June 27, 2009). Grant System Leads Cancer Researchers to Play It Safe. *New York Times*. Retrieved from <http://www.nytimes.com/2009/06/28/health/research/28cancer.html>
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Larivière, V., Haustein, S., & Börner, K. (2015). Long-Distance Interdisciplinarity Leads to Higher Scientific Impact. *Plos One*, 10(3), e0122565. doi:10.1371/journal.pone.0122565
- Lee, Y.-N., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy*, 44(3), 684-697. doi:<http://dx.doi.org/10.1016/j.respol.2014.10.007>
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71-87. doi:10.1287/orsc.2.1.71
- Martin, B. R. (2016). Editors' JIF-boosting stratagems – Which are appropriate and which not? *Research Policy*, 45(1), 1-7. doi:<http://dx.doi.org/10.1016/j.respol.2015.09.001>
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220-232. doi:10.1037/h0048850
- Merton, R. K. (1973). *The sociology of science*. Chicago, IL: University of Chicago Press.
- Monastersky, R. (2005). The number that's devouring science. *Chronicle of Higher Education*, 52(8), 14.
- Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press of Harvard University Press.
- NPR. (2013). Scientists win Nobel for work on how cells communicate. Retrieved from <http://www.npr.org/templates/story/story.php?storyId=230192033>
- Petsko, G. A. (2012). Goodbye, Columbus. *Genome Biology*, 13(5), 155. doi:10.1186/gb-2012-13-5-155
- Phiel, C. J., Zhang, F., Huang, E. Y., Guenther, M. G., Lazar, M. A., & Klein, P. S. (2001). Histone Deacetylase Is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen. *Journal of Biological Chemistry*, 276(39), 36734-36741. doi:10.1074/jbc.M101287200
- Planck, M. (1950). *Scientific autobiography, and other papers*. London, UK: Williams & Norgate.
- Rinia, E., Van Leeuwen, T., Bruins, E., Van Vuren, H., & Van Raan, A. (2002). Measuring knowledge transfer between fields of science. *Scientometrics*, 54(3), 347-362. doi:10.1023/A:1016078331752

- Romer, P. M. (1994). The Origins of Endogenous Growth. *The Journal of Economic Perspectives*, 8(1), 3-22. Retrieved from <http://www.jstor.org/stable/2138148>
- Rzhetsky, A., Foster, J. G., Foster, I. T., & Evans, J. A. (2015). Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences*, 112(47), 14569-14574. doi:10.1073/pnas.1509757112
- Schumpeter, J. A. (1939). *Business Cycles; A Theoretical, Historical, and Statistical Analysis of the Capitalist Process* (1st ed.). New York, NY; London, UK: McGraw-Hill Book Company, inc.
- Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*, 43, 73-85. doi:<http://dx.doi.org/10.1016/j.socnet.2015.02.006>
- Simon, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. G. Colodny (Ed.), *Mind and cosmos; essays in contemporary science and philosophy* (pp. 22-40). Pittsburgh: University of Pittsburgh Press.
- Simon, H. A., Langley, P. W., & Bradshaw, G. L. (1981). Scientific discovery as problem solving. *Synthese*, 47(1), 1-27.
- Simonton, D. K. (2004). *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*. Cambridge, UK; New York, NY: Cambridge University Press.
- StataCorp. (2016). nbreg. Retrieved from <http://www.stata.com/manuals13/rnbreg.pdf>
- Stent, G. (1972). Prematurity and uniqueness in scientific discovery. *Scientific American*, 227(6), 84-93.
- The National Academies. (2004). *Facilitating interdisciplinary research*. Washington DC: The National Academies Press.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157), 468-472. doi:10.1126/science.1240474
- Van Raan, A. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 467-472.
- Varian, H. R. (2009). The Economics of Combinatorial Innovation. *Guglielmo Marconi Lecture, Lisbon Council*. Retrieved from <https://www.youtube.com/watch?v=hqaA-fgdXEE>
- Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, 45(3), 707-723. doi:<http://dx.doi.org/10.1016/j.respol.2015.11.010>
- Walsh, D. (2013). Not Safe for Funding: The N.S.F. and the Economics of Science. *The New Yorker*. Retrieved from <http://www.newyorker.com/tech/elements/not-safe-for-funding-the-n-s-f-and-the-economics-of-science>
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851-872. doi:10.1007/s11192-012-0775-9
- Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and Impact: Distinct Effects of Variety, Balance, and Disparity. *Plos One*, 10(5), e0127298. doi:10.1371/journal.pone.0127298
- Weitzman, M. L. (1998). Recombinant Growth. *The Quarterly Journal of Economics*, 113(2), 331-360. Retrieved from <http://www.jstor.org/stable/2586906>
- Wyatt, H. (1975). Knowledge and prematurity: the journey from transformation to DNA. *Perspectives in Biology and Medicine*, 18(2), 149-156.
- Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *Plos One*, 10(8), e0135095. doi:10.1371/journal.pone.0135095
- Zhang, R., Wang, J., & Mei, Y. (2017). Search for evergreens in science: A functional data analysis. *Journal of Informetrics*, 11(3), 629-644. doi:<https://doi.org/10.1016/j.joi.2017.05.007>

Appendix I: Indicator illustration

We use the paper coauthored by Phiel et al. (2001), “Histone Deacetylase Is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen” published in the *Journal of Biological Chemistry* to illustrate our novelty measure. This paper cites 42 WoS-indexed journals. Of all possible journal pairs (861), 9 journal pairs are new, using the procedure described supra.

Pair	Journal 1	Journal 2	Novelty
1	GENE EXPRESSION	JOURNAL OF CLINICAL PSYCHIATRY	0.88
2	GENE EXPRESSION	NEUROPSYCHOPHARMACOLOGY	0.72
3	GENE EXPRESSION	ANNALS OF PHARMACOTHERAPY	0.78
4	GENE EXPRESSION	JOURNAL OF PHYSIOLOGY AND PHARMACOLOGY	0.52
5	GENE EXPRESSION	AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY	0.94
6	GENE EXPRESSION	KLINISCHE PADIATRIE	0.70
7	KLINISCHE PADIATRIE	NEUROPSYCHOPHARMACOLOGY	0.81
8	KLINISCHE PADIATRIE	JOURNAL OF PHYSIOLOGY AND PHARMACOLOGY	0.62
9	INTERNATIONAL IMMUNOLOGY	AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY	0.94

All journals in the “Journal 2” column (except *KLINISCHE PADIATRIE*) are in fields such as Psychiatry, Physiology, and Neurosciences, and these references are about valproic acid (VPA) in various settings such as epilepsy and bipolar disorder. All new journal pairs (except pair #6) are between these bodies of knowledge and three other journals. The first group of new journal pairs (pair #1-5) are with *GENE EXPRESSION*, and the reference is to the knowledge that histone deacetylase (HDAC) is a negative regulator of gene transcription. The second group (pair #7-8) is with *KLINISCHE PADIATRIE*, and the reference is to a study showing the effect of VPA on inhibiting proliferation of cell lines derived from human malignant gliomas, which make the connection between VPA and cancer (more precisely between the knowledge that VPA inhibits HDAC and that HDAC inhibitors prevents proliferation of cancer cells) more evident. Furthermore, the combination between *GENE EXPRESSION* and *KLINISCHE PADIATRIE* (pair #6) is also new. The third group (pair #9) is with *INTERNATIONAL IMMUNOLOGY*, and the

referenced paper demonstrates that lithium can alter gene expression. This pair is about combination of methods, i.e., invoking the procedure in the *INTERNATIONAL IMMUNOLOGY* paper for testing whether VPA activates Wnt-dependent gene expression and whether through the same mechanism as lithium.

In the next step we calculate the *cosine similarity* for each journal pair, based on their profile of co-cited journals in the preceding three years (i.e., 1998-2000). The co-citation matrix between all journals in the database is reported in the following table, based on the co-citation information between 1998 and 2000. One entry represents the number of times that two corresponding journals are co-cited in this time period. For example, *GENE EXPRESSION* has not been cited together with the *JOURNAL OF CLINICAL PSYCHIATRY* by any papers, but has been cited together with the *JOURNAL OF BIOLOGICAL CHEMISTRY* by 497 papers.

	J ₁	J ₂	J ₃	J ₄	J ₅	...	J _N
J ₁ GENE EXPRESSION	/	0	0	307	497
J ₂ JOURNAL OF CLINICAL PSYCHIATRY	0	/	1502	42	332
J ₃ NEUROPSYCHO-PHARMACOLOGY	0	1502	/	99	692
J ₄ NUCLEIC ACIDS RESEARCH	307	42	99	/	44713
J ₅ JOURNAL OF BIOLOGICAL CHEMISTRY	497	332	692	44713	/
...	/	...
J _N	/

Using the *cosine similarity* score, the ease of combining *GENE EXPRESSION* and the *JOURNAL OF CLINICAL PSYCHIATRY* is:

$$\begin{aligned} \text{COS}_{1,2} &= \frac{J_1 \cdot J_2}{\|J_1\| \|J_2\|} = \frac{\sum J_{1,i} \times J_{2,i}}{\sqrt{\sum J_{1,i}^2 \times \sum J_{2,i}^2}} \\ &= \frac{0 \times 1502 + 307 \times 42 + 497 \times 332 + \dots}{\sqrt{(0^2 + 0^2 + 307^2 + 497^2 + \dots) \times (0^2 + 1502^2 + 42^2 + 332^2 + \dots)}} = 0.12 \end{aligned}$$

Correspondingly, the difficulty of making this combination is: $1 - \text{COS}_{1,2} = 0.88$. We calculate the difficulty of all the other journal combinations following the same procedure. At the paper level, we sum up these difficulty scores and get a novelty measure, $\text{novelty} = 6.89$, which places the paper in the top 1% of novel papers in its field.

Appendix II: Variables, descriptive statistics, and additional regressions

Table A1. List of variables

Variable	Description
1 Novelty	Continuous combinatorial novelty score.
2 NOV CAT1	Novelty class dummy: 1 if non-novel, i.e., if a paper has no new journal combinations, and 0 otherwise.
3 NOV CAT2	Novelty class dummy: 1 if moderately novel, i.e., if a paper makes at least one new combination but has a novelty score lower than the top 1% of its subject category, and 0 otherwise.
4 NOV CAT3	Novelty class dummy: 1 if highly novel, i.e., if a paper has a novelty score among the top 1% of its subject category, and 0 otherwise.
5 Citations (15y)	Cumulative number of citations in a 15-year time window, i.e., 2001-2015.
6 Top 1% cited (15y)	Dummy, 1 if among the top 1% cited in the same subject category, based on 15-year citations.
7 Top 1% cited (3y)	Dummy, 1 if among the top 1% cited in the same subject category, based on 3-year citations.
8 Cited by big hits (10y)	Dummy: 1 if cited by a big hit published between 2001 and 2010. The citing big hits are identified as the top 1% highly cited article in the same WoS subject category and publication year, based on their cumulative citations till 2015. Only big hits published between 2001 and 2010 are identified, so that each citing article has at least six years to accumulate citations, for a reliable identification of citing big hit articles.
9 # citing fields	The number of WoS subject categories citing the focal paper. Coded as missing for papers without any citations.
10 Ratio of foreign field citations	The proportion of citations that are from papers sharing no common WoS subject categories, using a 15-year citation time window. Coded as missing for papers without any citations.
11 Max distance between the citing foreign and the home field	The maximum distance between a paper's home field and the foreign fields where it is cited (in a 15-year time window), where pairwise distance between two fields are calculated as $1 - \text{cosine similarity}$ between these two fields based on their co-citations in the preceding three year. Coded as missing for papers without any forward citations from foreign fields.
12 Top 1% cited in home fields (15y)	Dummy, 1 if among the top 1% cited in the same subject category, based on 15-year home field citations, i.e., citations received from subsequent papers sharing at least one common WoS subject categories.
13 Top 1% cited in foreign fields (15y)	Dummy, 1 if among the top 1% cited in the same subject category, based on 15-year foreign field citations, i.e., citations received from subsequent papers sharing no common WoS subject categories.
14 JIF	The Impact Factor (for year 2001) of the journal where the focal paper is published. It is missing for journal started after 1999.
15 International	Dummy: 1 if internationally co-authored, and 0 otherwise.
16 # authors	The number of authors.
17 # references	The number of references.

Table A2. Descriptive statistics and Spearman correlations

Variable	N	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1 Novelty	1038238	0.22	1.11	0	200.96																	
2 NOV CAT1	1038238	0.89	0.32	0	1	-1.00																
3 NOV CAT2	1038238	0.10	0.31	0	1	.94	-.95															
4 NOV CAT3	1038238	0.01	0.10	0	1	.31	-.28	-.03														
5 Citations (15y)	1038238	28.58	83.64	0	36792	.11	-.11	.10	.04													
6 Top 1% cited (15y)	1038238	0.01	0.10	0	1	.02	-.02	.02	.02	.17												
7 Top 1% cited (3y)	1038238	0.01	0.10	0	1	.01	-.01	.01	.01	.16	.45											
8 Cited by big hits (10y)	1038238	0.14	0.35	0	1	.05	-.05	.05	.03	.45	.23	.19										
9 # citing fields	979817	10.37	8.31	1	202	.15	-.15	.14	.06	.84	.16	.13	.38									
10 Ratio of foreign field citations	979817	0.47	0.29	0	1	.08	-.08	.08	.03	.06	.02	.01	.04	.25								
11 Max distance between citing foreign and home fields	881201	0.86	0.13	0.23	1	.12	-.12	.11	.05	.38	.12	.08	.21	.50	.19							
12 Top 1% cited in home fields (15y)	1038238	0.01	0.10	0	1	.01	-.01	.01	.01	.17	.66	.39	.22	.14	-.04	-.10						
13 Top 1% cited in foreign fields (15y)	1038238	0.01	0.10	0	1	.03	-.03	.02	.03	.17	.68	.37	.21	.16	.07	-.12	.36					
14 JIF	1006687	2.11	2.45	0	33.47	.04	-.05	.04	.01	.52	.08	.09	.23	.40	.06	-.02	.08	.07				
15 International	1038238	0.19	0.39	0	1	.01	-.01	.00	.01	.08	.02	.02	.04	.04	-.01	.01	.01	.01	.10			
16 # authors	1038238	4.14	6.20	1	743	.03	-.03	.03	.00	.19	.02	.04	.07	.17	.08	.02	.02	.02	.26	.20		
17 # references	1038238	28.78	17.98	2	631	.25	-.24	.22	.10	.39	.05	.06	.16	.32	.06	-.12	.05	.05	.32	.06	.01	

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection.

Table A3. Direct and indirect impact of novel research

	NOV CAT1	NOV CAT2	NOV CAT3
Probability of being top 1% cited (15-year)	0.90%	1.48%	3.14%
Average citations of citing papers ('01-'10)	28.11	35.81	39.51
Probability of a citing paper being top 1% cited ('01-'10)	1.36%	1.63%	2.00%
Probability of being cited by big hits (10-year)	13.34%	18.63%	23.95%

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection.

Table A4. Big hits in home fields

	Top 1% cited (1y) logit	Top 1% cited (2y) logit	Top 1% cited (3y) logit	Top 1% cited (4y) logit	Top 1% cited (5y) logit	Top 1% cited (6y) logit	Top 1% cited (7y) logit	Top 1% cited (8y) logit	Top 1% cited (9y) Logit	Top 1% cited (10y) logit	Top 1% cited (11y) logit	Top 1% cited (12y) logit	Top 1% cited (13y) logit	Top 1% cited (14y) logit	Top 1% cited (15y) logit
NOV CAT2	-0.28*** (0.04)	-0.36*** (0.04)	-0.28*** (0.03)	-0.26*** (0.03)	-0.23*** (0.03)	-0.22*** (0.03)	-0.21*** (0.03)	-0.19*** (0.03)	-0.15*** (0.03)	-0.12*** (0.03)	-0.12*** (0.03)	-0.11** (0.03)	-0.09** (0.03)	-0.10** (0.03)	-0.10** (0.03)
NOV CAT3	-0.41*** (0.12)	-0.51*** (0.10)	-0.35*** (0.08)	-0.20* (0.08)	-0.19* (0.08)	-0.17* (0.08)	-0.13+ (0.07)	-0.12 (0.07)	-0.06 (0.07)	-0.01 (0.07)	0.01 (0.07)	0.01 (0.07)	0.02 (0.07)	-0.02 (0.07)	0.01 (0.07)
International	0.17*** (0.03)	0.18*** (0.03)	0.15*** (0.02)	0.13*** (0.02)	0.12*** (0.02)	0.11*** (0.02)	0.11*** (0.02)	0.11*** (0.02)	0.10*** (0.02)	0.10*** (0.02)	0.08** (0.02)	0.09*** (0.02)	0.08** (0.02)	0.08** (0.02)	0.07** (0.02)
ln(# authors)	0.45*** (0.02)	0.68*** (0.02)	0.72*** (0.02)	0.71*** (0.02)	0.70*** (0.02)	0.68*** (0.02)	0.66*** (0.02)	0.63*** (0.02)	0.61*** (0.02)	0.60*** (0.02)	0.58*** (0.02)	0.57*** (0.02)	0.54*** (0.02)	0.52*** (0.02)	0.51*** (0.02)
ln(# refs)	0.86*** (0.02)	1.18*** (0.02)	1.25*** (0.02)	1.26*** (0.02)	1.26*** (0.02)	1.26*** (0.02)	1.26*** (0.02)	1.27*** (0.02)	1.25*** (0.02)	1.24*** (0.02)	1.22*** (0.02)	1.21*** (0.02)	1.21*** (0.02)	1.22*** (0.02)	1.22*** (0.02)
pubs.	660646	661244	661051	661413	661414	661414	661414	661389	661414	661414	661414	661389	661413	661413	661414
obs.	1034659	1036569	1036418	1037376	1037695	1037695	1037695	1037551	1037695	1037695	1037695	1037551	1037541	1037541	1037695
Pseudo R2	0.03	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04
Log lik	-38712	-46247	-50058	-51614	-52301	-52713	-53042	-53445	-53632	-54036	-54204	-54306	-54299	-54407	-54503

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Big hits are identified exclusively by home field citations, i.e., citations received from subsequent publications that share at least one common WoS subject category with the focal paper. Being a big hit in home fields using a t -year time window means being among the top 1% highly cited papers within the same WoS subject category, based on their home field citations in the t -year time window starting from 2001. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$.

Table A5. Big hits in foreign fields

	Top 1% cited (1y) logit	Top 1% cited (2y) Logit	Top 1% cited (3y) logit	Top 1% cited (4y) logit	Top 1% cited (5y) logit	Top 1% cited (6y) logit	Top 1% cited (7y) logit	Top 1% cited (8y) logit	Top 1% cited (9y) Logit	Top 1% cited (10y) logit	Top 1% cited (11y) logit	Top 1% cited (12y) logit	Top 1% cited (13y) logit	Top 1% cited (14y) logit	Top 1% cited (15y) logit
NOV CAT2	-0.14** (0.04)	0.01 (0.03)	0.07* (0.03)	0.16*** (0.03)	0.19*** (0.03)	0.24*** (0.03)	0.23*** (0.03)	0.26*** (0.03)	0.27*** (0.03)	0.29*** (0.03)	0.31*** (0.03)	0.30*** (0.03)	0.30*** (0.03)	0.31*** (0.03)	0.32*** (0.03)
NOV CAT3	-0.14 (0.11)	0.07 (0.08)	0.34*** (0.07)	0.39*** (0.06)	0.51*** (0.06)	0.51*** (0.06)	0.57*** (0.06)	0.61*** (0.06)	0.62*** (0.06)	0.63*** (0.06)	0.66*** (0.06)	0.68*** (0.06)	0.68*** (0.06)	0.67*** (0.06)	0.67*** (0.06)
International	0.24*** (0.03)	0.22*** (0.03)	0.14*** (0.02)	0.10*** (0.02)	0.07** (0.02)	0.06* (0.02)	0.05+ (0.02)	0.05+ (0.02)	0.04 (0.02)	0.04+ (0.02)	0.04+ (0.02)	0.04+ (0.02)	0.06* (0.02)	0.07** (0.02)	0.05* (0.02)
ln(# authors)	0.66*** (0.02)	0.75*** (0.02)	0.79*** (0.02)	0.77*** (0.02)	0.74*** (0.02)	0.73*** (0.02)	0.71*** (0.02)	0.67*** (0.02)	0.65*** (0.02)	0.61*** (0.02)	0.59*** (0.02)	0.57*** (0.02)	0.56*** (0.02)	0.54*** (0.02)	0.52*** (0.02)
ln(# refs)	0.98*** (0.03)	1.10*** (0.02)	1.18*** (0.02)	1.17*** (0.02)	1.21*** (0.02)	1.17*** (0.02)	1.17*** (0.02)	1.15*** (0.02)	1.17*** (0.02)	1.16*** (0.02)	1.15*** (0.02)	1.16*** (0.02)	1.15*** (0.02)	1.15*** (0.02)	1.14*** (0.02)
pubs.	658492	661120	661388	661414	661413	661413	661414	661302	661414	661414	661413	661157	661413	661413	661414
obs.	1027507	1036705	1037397	1037695	1037541	1037541	1037695	1037244	1037695	1037695	1037541	1037244	1037541	1037541	1037695
Pseudo R2	0.06	0.05	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Log lik	-33556	-44923	-49377	-51076	-51964	-52644	-53079	-53315	-53501	-53749	-53984	-54102	-54129	-54200	-54324

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Big hits are identified exclusively by foreign field citations, i.e., citations received from subsequent publications that share no common WoS subject categories with the focal paper. Being a big hit in foreign fields using a t -year time window means being among the top 1% highly cited papers within the same WoS subject category, based on their foreign field citations in the t -year time window starting from 2001. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$.

Table A6. Big hits: novelty interact with JIF

	Top 1% cited (1y) logit	Top 1% cited (2y) Logit	Top 1% cited (3y) logit	Top 1% cited (4y) logit	Top 1% cited (5y) logit	Top 1% cited (6y) logit	Top 1% cited (7y) logit	Top 1% cited (8y) logit	Top 1% cited (9y) Logit	Top 1% cited (10y) logit	Top 1% cited (11y) logit	Top 1% cited (12y) logit	Top 1% cited (13y) logit	Top 1% cited (14y) logit	Top 1% cited (15y) logit
NOV CAT2	-0.14* (0.06)	-0.04 (0.05)	0.08+ (0.05)	0.13** (0.05)	0.21*** (0.05)	0.22*** (0.04)	0.25*** (0.04)	0.27*** (0.04)	0.30*** (0.04)	0.32*** (0.04)	0.32*** (0.04)	0.30*** (0.04)	0.32*** (0.04)	0.32*** (0.04)	0.33*** (0.04)
NOV CAT3	-0.01 (0.14)	0.19 (0.12)	0.51*** (0.10)	0.71*** (0.09)	0.75*** (0.09)	0.76*** (0.09)	0.79*** (0.09)	0.79*** (0.09)	0.82*** (0.08)	0.81*** (0.08)	0.79*** (0.08)	0.82*** (0.08)	0.84*** (0.08)	0.86*** (0.08)	0.88*** (0.08)
JIF TOP10%	1.60*** (0.03)	1.95*** (0.03)	2.06*** (0.03)	2.04*** (0.03)	2.05*** (0.03)	1.98*** (0.03)	1.95*** (0.03)	1.91*** (0.03)	1.88*** (0.03)	1.85*** (0.03)	1.83*** (0.02)	1.79*** (0.02)	1.77*** (0.02)	1.73*** (0.02)	1.72*** (0.02)
JIFTOP *	-0.15+ (0.08)	-0.18** (0.07)	-0.21** (0.06)	-0.23*** (0.06)	-0.32*** (0.06)	-0.30*** (0.06)	-0.30*** (0.06)	-0.28*** (0.06)	-0.31*** (0.06)	-0.32*** (0.06)	-0.32*** (0.06)	-0.30*** (0.06)	-0.32*** (0.06)	-0.30*** (0.06)	-0.31*** (0.06)
NOVCAT2	-0.33 (0.21)	-0.57** (0.17)	-0.63*** (0.14)	-0.67*** (0.13)	-0.71*** (0.13)	-0.62*** (0.13)	-0.66*** (0.13)	-0.62*** (0.13)	-0.60*** (0.12)	-0.57*** (0.12)	-0.55*** (0.12)	-0.60*** (0.12)	-0.60*** (0.12)	-0.60*** (0.12)	-0.60*** (0.12)
JIFTOP *	0.17*** (0.03)	0.14*** (0.03)	0.09*** (0.02)	0.08** (0.02)	0.06** (0.02)	0.06** (0.02)	0.06** (0.02)	0.05* (0.02)	0.06* (0.02)	0.06* (0.02)	0.05+ (0.02)	0.04 (0.02)	0.03 (0.02)	0.04+ (0.02)	0.04+ (0.02)
International	0.48*** (0.02)	0.62*** (0.02)	0.68*** (0.02)	0.65*** (0.02)	0.63*** (0.02)	0.60*** (0.02)	0.59*** (0.02)	0.56*** (0.02)	0.53*** (0.02)	0.51*** (0.02)	0.48*** (0.02)	0.46*** (0.02)	0.44*** (0.02)	0.42*** (0.02)	0.41*** (0.02)
ln(# authors)	0.78*** (0.03)	0.97*** (0.02)	1.01*** (0.02)	1.00*** (0.02)	1.01*** (0.02)	1.01*** (0.02)	1.00*** (0.02)	0.98*** (0.02)	0.97*** (0.02)	0.97*** (0.02)	0.97*** (0.02)	0.97*** (0.02)	0.97*** (0.02)	0.97*** (0.02)	0.96*** (0.02)
ln(# refs)	641480	642544	642674	642674	642674	642535	642674	642674	642674	642674	642674	642673	642674	642674	642674
pubs.	1002498	1006168	1007146	1007146	1006679	1006920	1007146	1007146	1007146	1007146	1007146	1006992	1007146	1007146	1007146
obs.	0.09	0.12	0.13	0.13	0.13	0.12	0.12	0.12	0.11	0.11	0.11	0.10	0.10	0.10	0.10
Pseudo R2	-37375	-44407	-45768	-46986	-47461	-48113	-48286	-48926	-49135	-49317	-49616	-49754	-49920	-50100	-50260
Log lik															

Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection. Big hits are identified based on all citations (i.e., both home and foreign field citations). Being a big hit in a t -year time window means being among the top 1% highly cited papers within the same WoS subject category and publication year, based on their citations in the t -year time window starting from 2001. JIFTOP is a dummy variable, 1 if the impact factor of the journal in which the paper is published is among the top 10% in its field. Field (WoS subject category) fixed effects incorporated. Robust standard errors in parentheses. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$.

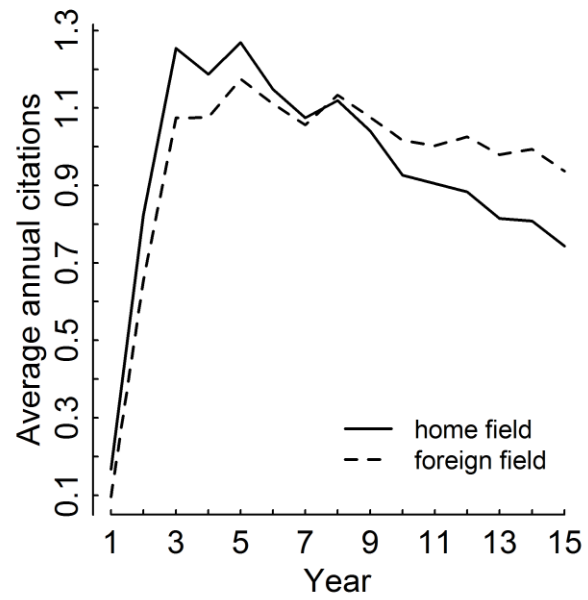


Figure A1. Average annual citations. The average number of annual citations over time, separating home and foreign fields. Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection.

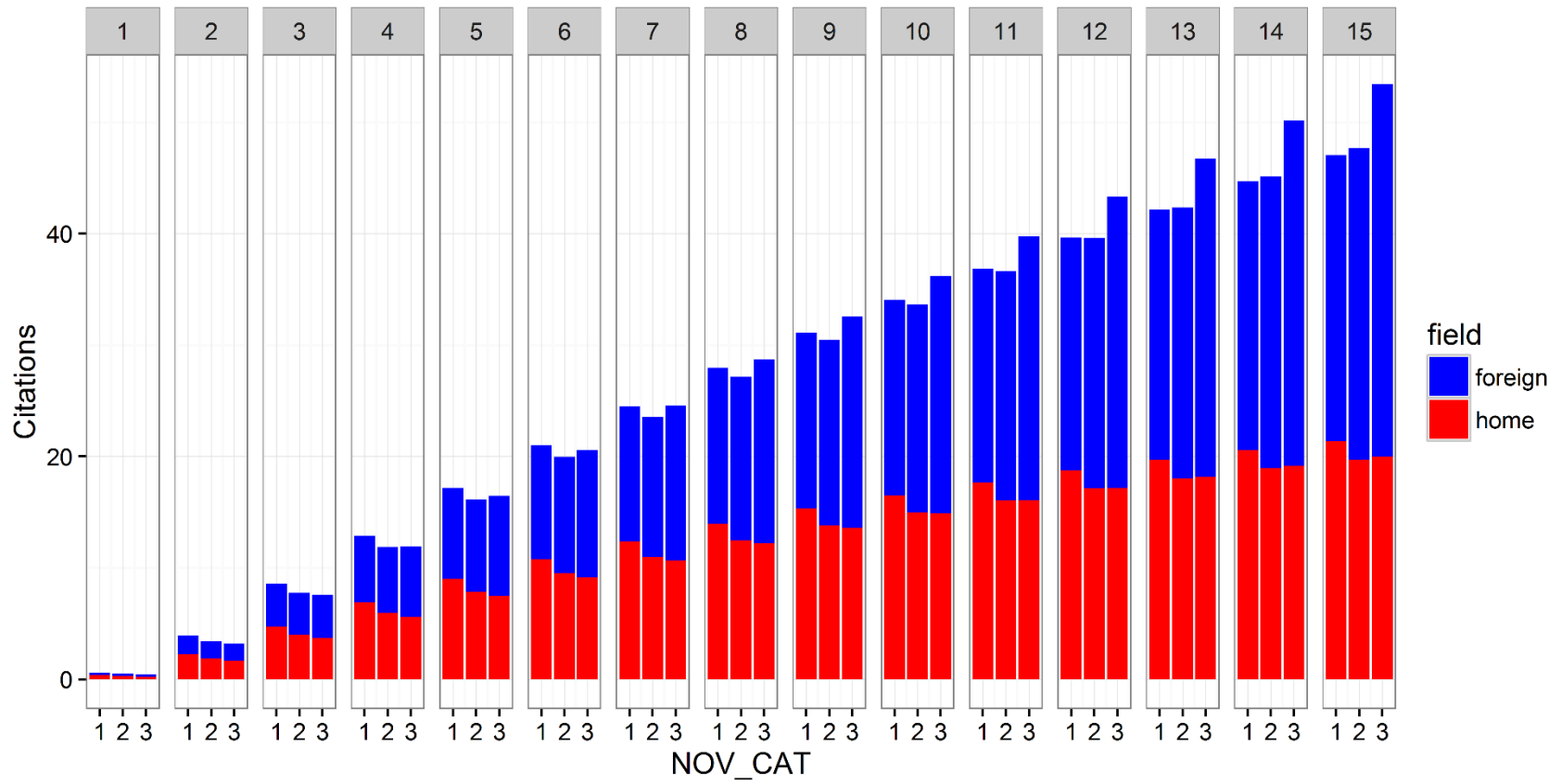


Figure A2. Estimated cumulative number of citations by novelty classes and time windows, separating home and foreign field citations. Estimates are based on a set of Poisson regressions. Using negative binomial models yields similar results. Data consist of all WoS articles published in 2001 and are sourced from Web of Science Core Collection.

Appendix III: Robustness analysis

Our major results can be summarized as follows: novel papers have (1) a higher dispersion in citations, (2) a higher chance of being a big hit in the long term, (3) a higher chance of being cited by big hits, (4) a broader impact across scientific fields, (5) a larger share of citations from foreign fields, (6) a greater transdisciplinary impact reaching fields more distant from their home field, (7) not a higher chance of being highly cited in home field, (8) a higher chance of being highly cited in foreign field, (9) a lower chance of being a big hit in the short term, and (10) are published in journals with lower Impact Factors. In this section we discuss the robustness of these ten findings. Results are not reported but are available upon request to the authors.

III.1. Scientific field heterogeneity

Although we control for scientific field effects, and some of our measures for novelty (i.e., highly novel) and impact (e.g., big hits) are field specific, it is nevertheless possible that the relationship between novelty and impact is field specific. To test this possibility, we examine whether our main findings are robust across scientific fields, replicating the whole set of analyses separately by field. We use four groups: AH (Arts and Humanities), LS (Life Sciences), PSE (Physical Sciences and Engineering), and SS (Social Sciences). We also distinguish between LS2 (Medicine) and the rest (LS1) within LS and distinguish between PSE2 (Computer Sciences; Engineering) and the rest (PSE1) within PSE. Both LS2 and PSE2 are the more applied counterparts of LS1 and PSE1. In total, we run separate regressions for the following eight groups: AH, LS, LS1, LS2, PSE, PSE1, PSE2, and SS. While our findings are robust for LS and PSE fields, they are less robust for AH and SS fields. Specifically, findings, (1) higher dispersion, (9) fewer short-term big hits, and (10) lower JIF for novel papers, are not robust for AH, and finding (1) higher dispersion is not robust for SS. This may suggest that the relationship between novelty and impact varies across fields and our findings hold most specifically for hard sciences and engineering. However, it may also be an artifact of data limitations. The WoS coverage for humanities and social sciences is much smaller than hard sciences, so we have much fewer observations for AH and SS than for the hard sciences. More importantly, because of this coverage issue, using the WoS data does not capture sufficiently all the publishing and

referencing transactions in humanities and social sciences (Hicks, 2004), which is required for constructing measures of novelty and impact.

In addition, results reported in this paper are based on a dataset consisting of 661,643 unique publications and 1,038,238 observations, where papers with multiple WoS subject categories are counted multiple times. This might give those papers with multiple subject categories higher weight in influencing the results. Additionally, we run two sets of robustness checks for all our results: (1) analyzing papers with only one subject category and (2) reassigning papers with multiple subject categories and papers in the subject category of “Multidisciplinary Sciences” to one single subject category, the majority one of their references. All our findings are robust in these two sets of robustness analyses.

III.2. Novelty measure variations

First, we test whether our findings hold when we use alternative formulas for our novelty measure. As explained in section 3.1, our novelty measure is essentially the number of new referenced journal pairs, weighted by the distance between the newly paired journals (i.e., *weighted sum*). Alternatively we can use the novelty score of the most distinct pair among all the new pairs (i.e., *max score*) or the proportion of all journal pairs that are new, weighted by the distance (i.e., *weighted share*).

The maximum novelty score,

$$Novelty_{max\ score} = \max_{J_i - J_j \text{ pair is new}} (1 - COS_{i,j})$$

focuses exclusively on the novelty score of the most distant new journal pair instead of mixing both the number and the difficulty of new journal pairs. We find using the *max score* approach that all our findings hold, except finding 7, specifically, while *max score CAT2* also has a significantly negative effect on being highly cited in the home field, *max score CAT3* has a small but significantly ($p = 0.08$) positive effect. One possible explanation is that papers with a small number of very distant new combinations but at same time a large number of conventional combinations within the home field are likely to score high on the *max score* measure, and the

large number of home field combinations makes the paper more likely to be accepted by the home field.

Because of the concern that our *weighted sum* measure of novelty might be over-dependent on the number of references in the focal paper, which including the number of references in the regression analysis may not sufficiently control for, we also inspect the *weighted share* approach, which essentially normalizes our novelty measure for the number of journal pairs. It can also be viewed as a reversed and weighted *network density*¹¹ measure with the number of missing ties (i.e., new ties which did not exist in previous years, weighted by the distance) in the numerator and the size of the network (i.e., the number of all journal pairs regardless new or not) in the denominator. Specifically,

$$Novelty_{weighted\ share} = \frac{1}{n} \sum_{J_i - J_j\ pair\ is\ new} (1 - COS_{i,j})$$

where n is the number of all co-cited journal pairs. All findings hold using this *weighted share* approach. The *weighted share* measure imposes a penalty on papers with a large number of referenced journals, in other words, papers with a larger network of referenced journals. We believe that network size is also an important aspect of combinatorial novelty and therefore should not be eliminated from the analysis. Specifically, a larger network making the same number of new connections should not be evaluated as less novel than a smaller network. Furthermore, the *weighted share* measure is very sensitive when a paper has very limited number of referenced journals. Therefore, we prefer our *weight sum* measure to the *weighted share* approach.

Second, our novelty indicator used in the analysis excluded the 50% least cited journals and required reuse of the new combination in the next three years. To check the sensitivity of the analysis to these choices, we replicate the analysis without these restrictions. In addition, although most new journals are already excluded in the analysis because they are typically among the lower cited journals, we also checked the results excluding new journals from the calculation

¹¹ Network density is the proportion of all possible ties that are actually present in the network.

of *novelty*. Moreover, Boyack and Klavans (2014) warned that the *atypicality* measure in Uzzi et al. (2013) is confounded with citing star journals, such as *Science* and *Nature*. Our novelty measure is unlikely to be confounded with citing star journals, which are so highly cited that combining them with any other journals is very unlikely to be new. Nevertheless, we check results excluding the top 10% highly cited journals, as well as multidisciplinary journals. In summary, we test robustness of our findings using the following variations of the *novelty* measure: (a) no restrictions, (b) only exclude the 50% least cited journals, (c) only require reuse in 3 years, (d) only exclude new journals, (e) exclude the 50% least cited journals, require reuse in 3 years, and exclude the 10% top cited journals, and (f) exclude the 50% least cited journals, require reuse in 3 years, and exclude the 10% top cited and multidisciplinary journals. Our findings are robust using all these alternative specifications.

Third, to accommodate the skewness of the indicator, we created three novelty categories. When we duplicate the results using the natural logarithm transformed continuous novelty score in the regression, all results are robust.

Fourth, we classified highly novel papers as the top 1% novel papers in the same subject category. We also use different thresholds for the categorization, specifically, classifying the top 5% or 0.1% papers as highly novel. All results are robust, and the changes in effect size are consistent with what we would expect from comparing the coefficients on highly and moderately novel classes in our reported results. For example, compared with our reported results for classifying the top 1% papers as highly novel, highly novel would have a larger effect on the likelihood of big hits when classifying top 0.1% papers as highly novel but a smaller effect when classifying top 5% papers as highly novel.

III.3. Novelty versus atypicality

We compare our novelty indicator with the Uzzi et al. (2013) measure of *atypicality*. For a scientific publication, Uzzi et al. (2013) calculated the commonness score for all its referenced journal pairs, where the commonness score is a z -score, (i.e. the number of observed co-citations between the pair – the number of expected co-citations) / standard deviation of the co-citations) and labelled the lowest 10th percentile of this series of commonness scores as their indicator of novelty. To distinguish it from our measure of novelty, we label their indicator as *atypicality*.

Lee et al. (2015) used an adapted version in their study. Specifically, they used the ratio between the number of observed co-citations and the number of expected co-citations as the commonness score of a journal pair, instead of the z -score. In addition, they take the natural logarithm of the commonness score and add a negative sign, obtaining a roughly normally distributed variable positively associated with *atypicality*, while the commonness score is negatively associated with *atypicality*. Here we follow this adapted version to calculate the Uzzi et al. (2013) indicator for our sampled papers. To make the comparison, we categorize the continuous *atypicality* indicator into three classes with the same proportion of papers in each of the three classes, specifically, *ATP CAT* = 3 if top 1% in the same WoS subject category and publication year, 2 if below top 1% but above top 10%, and 1 for all others. In addition, we also compare the continuous versions of both *atypicality* and our *novelty* scores. Specifically, we assess results of six sets of regressions using the following independent variables: (a) *novelty* categories (*NOV CAT*), i.e., main results reported in the text, (b) *atypicality* categories (*ATP CAT*), (c) *NOV CAT* and *ATP CAT* together, (d) *novelty* score, (e) *atypicality* score, (f) *novelty* and *atypicality* scores together.

We focus on two criteria for assessing the results: the size of the effect and the consistency of the results across these six settings. When both categorical *novelty* and *atypicality* are in the regression, i.e., in setting (c), all effects of *novelty* (i.e., *NOV CAT*2 and *NOV CAT*3 on findings 1-10) remain consistent. *Atypicality* classes, compared with *novelty* classes, have smaller effects on (1) higher dispersion, (3) higher indirect impact, and (9) fewer short-term big hits; larger effect on (2) more long-term big hits and (4) broader impact, (5) higher share of foreign field citations, (6) greater transdisciplinary impact reaching more distant fields, (7) fewer big hits in home field, and (8) more big hits in foreign field; and reversed effect on (10) low JIF¹². When having both continuous *novelty* and *atypicality* scores in the regressions, i.e., in setting (f), again all the effects of *novelty* remain consistent, while the *atypicality* score, compared with the *novelty* score, has reversed effect on (9) fewer short-term big hits and (10) lower JIF; smaller effect on (1) higher dispersion, (3) higher indirect impact, and (6) greater transdisciplinary impact reaching more distant fields; and larger effect on (2) more long-term big hits, (4) broader impact, (5)

¹² Specifically, *ATP CAT*2 has a positive and significant effect, while *ATP CAT*3 has a positive but insignificant effect on JIF.

higher share of foreign field citations, and (8) more big hits in foreign field. Based on the comparison of effect size, it seems that our *novelty* indicator better captures the “high risk” nature of novel research than does the *atypicality score*. In terms of “high gain,” *atypicality* better captures the higher direct impact and broader impact, while *novelty* better captures the indirect impact of stimulating follow-on breakthroughs.

In terms of consistency across these six settings, our *novelty* measure has considerably more reliable and robust behavior than *atypicality*. While effects of our *novelty* measure are robust across all six settings, the results of *atypicality* are not consistent across settings. For example, *ATP CAT3* has a significantly negative effect on short-term big hits, but *atypicality* score has a significantly positive effect; and *atypicality* has a significantly negative effect on JIF in setting (b) but a positive effect in other settings. Overall, we conclude that our *novelty* indicator behaves more reliably, captures more the “high risk” nature of novel research, and measures aspects of research not captured by the *atypicality* measure.