STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

Novelty in Science: The impact of French physicists' novel articles

Jacques Mairesse* and Michele Pezzoni**

*jacques.mairesse@ensae.fr
Maastricht University (unu-merit), CREST-ENSAE, EHESS and NBER, 18, boulevard Gabriel Péri, 92245, Malakoff cedex (France)

**michele.pezzoni@unice.fr
Université Côte d'Azur, CNRS, GREDEG, and HCERES-OST, Campus Azur, 250 Rue Albert Einstein, CS 10269 06905 SOPHIA ANTIPOLIS CEDEX (France)

Introduction

A scientific article requires a certain degree of novelty to be published in a scientific journal with respect to the state of the art literature. However, some articles are more novel than others and introduce breakthrough ideas. The current literature is still debating on the definition of novel articles introducing novel ideas and on how to measure their impact. Moreover, little attention has been devoted to the impact of follower articles reusing novel ideas.

In this paper we investigate two issues. First, what is the impact of publishing an article introducing a novel idea, i.e. a novel article, in terms of forward citations received and quality of the journal where it is published. Second, what is the impact of an article reusing a novel idea after its introduction, i.e. a follower article. We conduct our study within the domain of physics by analysing the population of French researchers active in the year 2005.

In innovation studies, novel ideas are considered as the result of the recombination of existing pieces of knowledge in an original and unprecedented way (Schumpeter, 1939). These studies, aiming to identify novel (technological) ideas use patent information to operationalize the concept. Precisely, the technological classes reported on patent documents are considered as the pieces of technological knowledge recombined. An unprecedented combination of technological classes appearing in a patent document is considered as technological novelty (Fleming, 2001; Verhoeven et al., 2016).

Similarly, within the domain of science, publication data allow researchers to identify the existing pieces of knowledge and to trace the appearance of novel combinations. The most common approach to operationalize the novelty concept in science is to consider the journals referenced in scientific articles as pieces of knowledge and to consider novelty as an unprecedented combination of these journals (Ayoubi et al., 2017; Uzzi et al., 2013; Wang et al., 2017).

In this paper we identify novelty as a novel combination of referenced journals appearing for the first time in an article. We define a novel article as an article reporting a novel combination of referenced journals in its reference list.

Not all the novelty defined as an unprecedented journal combination is successful and favours the progress of the discipline. A novel journal combination that is repeatedly used after its first appearance, indicates a successful novel idea. Differently, a novel journal combination

that is abandoned after its first introduction, might indicate a trivial journal combination, a not promising research line, or simply a measurement error given by a mistake in reporting the journal name in the article references. In this paper we consider only novelty that is used after its first appearance neglecting all the trivial or abandoned combinations.
Although some novel articles might incorporate completely the essence of a novel idea, we expect that large part of the novel ideas needs to be refined with further developments since their appearance. This refinement process takes time and is documented in additional articles that are expected to have an impact on the discipline, although they are not identified as novel articles. For this reason, in this paper we extend the analysis of the impact of novelty to its further developments by tracing the impact of all the follower articles including the same novel idea.
We find that novel and follower articles published by French scientists receive the same number of citations in the five years after publication as non-novel articles and that they are published in lower impact factor journals. Moreover, when we distinguish short and long run citations, we find that novel and follower articles tend to receive more citations than non-novel articles in the long run, i.e., 3/4 years after their publication.
We conduct our analysis in three steps. First, we identify the novel articles within the universe of articles published in the domain of physics. To do so, we rely on the Web of Science core collection bibliometric dataset (WOS). Second, we identify the novel articles published by French physicists by matching the article authors with a list of names of physicists active in 2005 in all French universities and at CNRS. Finally, we assess the impact of novel and follower articles published by French physicists with respect to non-novel articles.

Novelty, novel articles, and follower articles in the universe of physics
To operationalize the concept of novelty, we start from the assumption that the journals listed in the article references represent the pieces of knowledge mobilized by the scientist to write the article. When two journals are listed together in the references of an article for the first time in the history of the discipline, they represent a novel combination. We apply our definition of novelty to the domain of physics. To do so, we start by defining the universe of articles in physics.
We consider as universe all the articles published during the period 2005-2009 in the 273 journals in physics with an impact factor larger than 0.5. We include in our analysis a total of 473,314 articles. These articles list 5,226 distinct referenced journals.
Starting from the complete list of referenced journals, we construct all the possible referenced journal combinations for each of the 473,314 articles. These combinations include both novel and non-novel combinations. A combination is novel if it does not appear in the references of another scientific article published in the preceding years. As an illustrative example, consider the article A published in year t that references the journals J1, J2, and J3 in its bibliography. All the possible combinations of the referenced journals are J1-J2, J1-J3, and J2-J3. If the combination J1-J2 does not appear before t in another article, then J1-J2 is a novel combination.
In our dataset, WOS publication data are available since 2000. For a journal combination referenced in an article published in year t, we observe the preceding publication history from 2000 to t-1. To identify novel combinations, we define a buffer period of an arbitrary length of 5 years, from 2000 to 2004, that represents the minimum period needed to claim that the journal combination is novel. For instance, in evaluating the novelty of the combination J1-J2 included in an article published in t=2005, we screen all the articles published in the period 2000-2004 to verify if the same combination J1-J2 have previously appeared. If not, then J1-J2 is a novel combination.

We consider only successful novelty namely, the novelty that is used in other scientific articles after its first appearance. Thus, we restrict our definition of novelty to the journal combinations that are used at least in 10 distinct articles in the five years after their first appearance.
We define as novel articles all the articles that include at least one novel combination and that are published the year when the novel combination appears. The universe of physics considered includes 473,314 articles, out of which 20,812 (4.4%) are novel articles.
Rarely a novel scientific idea can be associated to a unique article. Therefore, we define the follower articles as the articles using the novel combination up to four years after its first appearance. We identify 46,650 (9.8%) follower articles in the universe of physics.
Table 1 shows an example of a novel combination of two journals, ACTA PHYSICA POLONICA A and MATERIALS SCIENCE AND ENGINEERING A, that appeared for the first time in 2006. The journal ACTA PHYSICA POLONICA A is a highly-reputed journal in physics established by the Polish Physical society in 1920 that regularly publishes 12 issues per year. The journal MATERIALS SCIENCE AND ENGINEERING A is a journal publishing theoretical and experimental studies on loading-bearing capacity of the materials. The first issue has been published in 1988. The combination of ACTA PHYS. and MAT. SCI. satisfies all the criteria defined above to identify a novel combination namely, the combination does not appear in the buffer period 2000-2005 and is used in at least 10 articles after its first appearance. That novel combination is included in 4 novel articles published in 2006. Moreover, the journal combination is used in other 18 articles in the 4 years following its first appearance.

Table 1 Example of a novel combination

J1	J2	Appeared for the first time	year	N. of articles using the combination
ACTA PHYS.	*MAT. SCI.*	2006	2006	4
ACTA PHYS.	*MAT. SCI.*	2006	2007	1
ACTA PHYS.	*MAT. SCI.*	2006	2008	2
ACTA PHYS.	*MAT. SCI.*	2006	2009	8
ACTA PHYS.	*MAT. SCI.*	2006	2010	7

Among the 4 novel articles published in 2006, one, in the journal PHYSICAL REVIEW B, is authored by two French researchers. This article is cited 12 times in the following 5 years. The impact factor of PHYSICAL REVIEW B in 2006 equals 10.3, i.e. on average an article published on the journal receives 10.3 citation in the next 5 years after the publication. The novel article published in PHYSICAL REVIEW B and authored by the two French researchers is part of our study sample described in the next section.

Data and variables

The study sample of French physicists
This section describes our study sample, that is the collection of all the articles, novel and non-novel, published in the 273 physics journals including at least one French physicist among the authors. We identify French physicists relying on the list of names of active researchers in the year 2005 in the two main public French research organizations, namely universities and CNRS. For each researcher, we collect biographical characteristics such as her gender, rank, age, and bibliometric data. Our analysis considers the researchers' publications within the period 2005-2009.

We end up with a sample of 1090 physicists publishing 9042 articles during the period 2005-2009. We include in our analysis 564 university professors and 526 CNRS researchers. These scientists have at least one publication between 2005 and 2009. Out of 1090 scientists, 123 have at least one novel article, while 353 have at least one follower article that uses a combination that was introduced in the preceding 4 years. 692 French researchers do not publish neither novel articles nor follower articles during our study period. Among the 9,027 articles published by French physicists, 1.5% articles are novel articles, 2.8% are follower articles using a novel combination in the 2 years after the combination was introduced, and 3.1% are follower articles using that novel combination after 3 or 4 years its introduction. We run our regression exercise on a study sample at the researcher-article level, including 11,753 observations.

Dependent variables

We measure the impact of an article by counting its forward citations in the five years after its publication (Citations) and by recording the impact factor of the journal where the article is published (IF). The two impact measures, often used as substitutes, are conceptually different. Citations are counted after the article publication and reflects both the intrinsic quality of the article, including its novelty, and the reputation of the authors within the scientific community (Azoulay et al., 2010; Merton, 1968). On the contrary, the quality of the journal where the article is published, as represented by the impact factor, is the result of a process that takes place before the article is made publicly available to the scientific community involving editors and authors. Editors might have a bias against novelty denying the publication of novel articles due to the intrinsic risk of publishing unexplored novel ideas (Wang et al., 2017). In this paper we argue that the high number of citations received by a novel article might not correspond to its publication on a high quality journal.

Explanatory variables

Our main explanatory variable is the dummy *Novel article* that equals one if the article includes at least one novel journal combination, zero otherwise. Moreover, we define two dummies identifying follower articles that use a novel combination 1-2 or 3-4 years after the first use, respectively. Precisely, the dummy *Follower article 1-2yrs* equals one if the follower article uses the novel combination 1 or 2 years after its first appearance. Finally, *Follower article 3-4yrs* equals one if the follower article uses a novel combination 3 or 4 years after its first appearance.

Novel combination characteristics

We characterize the novel combination of referenced journals reported in a novel article. The novel articles authored by a French physicist included in our study sample might not be the only one reporting the novel combination when we consider the whole universe of physics. The novel combination characteristic variables are constructed according to the characteristics of the novel articles reporting the novel combination in the whole universe of physics, whether they are authored by a French physicist or not. Precisely, we count the number of novel articles referencing the novel combination (*Universe n. articles*). We count the number of distinct authors of those articles (*Universe n. authors*), and the presence of at least one US author (*Universe US author*). All these variables are measured conditional on observing a novel article. Therefore, in the regression exercise, they are included as interactions with the dummy *Novel article*.

Other controls
We consider also controls for: article characteristics, scientist characteristics, and calendar year and sub-discipline controls.

Table 2 and 3 show the descriptive statistics.

Table 2 Descriptive statistics.

	Obs.	Mean	sd	min	max
Dependent variables					
Citations	11753	13.595	45.501	0	1536
IF	11753	8.214	5.76	0.503	60.683
Independet variables					
Novel article	11753	0.018	0.133	0	1
Follower article 1-2 yrs	11753	0.03	0.17	0	1
Follower article 3-4 yrs	11753	0.022	0.145	0	1
Article characteristics					
Article n. of authors	11753	6.33	8.653	1	458
Article US affiliation	11753	0.099	0.299	0	1
N. of references	11753	30.61	35.37	2	1082
Scientist characteristics					
Female scientist	11753	0.109	0.312	0	1
CNRS scientist	11753	0.582	0.493	0	1
Scientist Age	11753	47.989	9.187	33	70
Scientist's publication stock (log)	11753	3.096	0.732	0	4.875
Scientist's average citations (log)	11753	1.47	0.877	-1.609	5.22
Scientist's average IF (log)	11753	0.7	0.682	-1.609	3.063
Condensed matter	11753	0.68	0.466	0	1

Table 3 Descriptive statistics of the novel characteristics

	Obs.	mean	sd	min	max
Universe n. Articles	213	2.748	1.361	1	8
Universe n. authors	213	4.491	2.264	1	12
Universe US author	213	0.432	0.497	0	1

Results
Table 4 shows the results of OLS estimations. Columns from 1 to 4 consider citations as dependent variable[1], whereas columns from 5 to 8 consider the impact factor as dependent variable. We include progressively the controls namely, article characteristics, scientist characteristics, and scientist fixed effects. The controls calendar year and sub-discipline are common to all the model specifications. Table 5 reports our main model specifications, namely columns 3-4 and 7-8 of Table 4, adding the novel combination characteristics.
We find that novel articles (*Novel article*) obtain the same number of citations as non-novel articles according to our preferred estimation reported in Column 4 of Table 4. Moreover,

[1] In these columns we use the approximation log(1+Citations) as dependent variable.

novel articles are published in journals with lower impact factor with respect to non-novel articles. Precisely, the coefficient of the variable *Novel article* equals -0.088 (-8.8%) and is statistically significant in our preferred specification reported in Column 8, Table 4. The presence of one additional author introducing the novel idea (*Novel article*Universe n. authors*) increases the number of citations received by 6.6% (Column 2, Table 5).
Interestingly, the follower articles (*Follower article 1-2yrs* and *Follower article 3-4yrs*) receive the same number of citations as the non-novel articles, while the impact factor of the journals where they are published remains significantly lower than the non-novel articles. For the novel articles, the citation premium is higher if the novel idea is introduced by a higher number of authors (Column 2, Table 5).
In Table 6 we report a series of estimations where we consider five alternative dependent variables counting the citations received each year since the article publication. Specifically, we calculate the citations received by each article the year when it is published (t=0), one year after its publication (t=1), two years after its publication (t=2), three years after its publication (t=3), and four years after its publication (t=4). We find that novel and follower articles are more cited in the long run (4 years after they are published) than non-novel articles (Column 9, Table 6).

Table 4 Regressions (Standard errors clustered around the researcher)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	log(Citations)	log(Citations)	log(Citations)	log(Citations)	log(JIF)	log(JIF)	log(JIF)	log(JIF)
Novel article	0.37***	0.064	0.090	0.059	-0.028	-0.19***	-0.16***	-0.088**
Follower article 1-2yrs	0.33***	0.071	0.092	0.087	-0.14***	-0.25***	-0.22***	-0.16***
Follower article 3-4yrs	0.10	-0.059	-0.025	-0.0030	-0.17***	-0.25***	-0.22***	-0.17***
Article n. of authors		0.013***	0.011***	0.0098***		-0.0016	-0.00061	0.00020
Article US affiliation		0.22***	0.21***	0.22***		0.13***	0.090***	0.086***
N. of references (log)		0.57***	0.54***	0.55***		0.27***	0.24***	0.23***
Female scientist			-0.040				-0.021	
CNRS scientist			0.048				0.028	
Scientist Age			-0.0065***	0.084***			-0.0033***	0.048**
Scientist's publication stock (log)			0.21***	-0.47***			0.29***	-0.17
Scientist's average citations (log)			0.24***	-0.40***			-0.021	0.11***
Scientist's average IF (log)			-0.027	0.092			0.39***	-0.55***
Condensed matter	0.044	0.034	0.047		0.095***	0.088***	0.062***	
Constant	1.97***	0.086	-0.47***	-1.84***	1.76***	0.90***	0.11	-0.45
Observations	11,753	11,753	11,753	11,753	11,753	11,753	11,753	11,753
R-squared	0.014	0.134	0.173	0.129	0.013	0.065	0.121	0.052
Dummy year	yes	yes	yes	yes	yes	yes	yes	yes
Scientist fixed effects	no	no	no	yes	no	no	no	yes

Table 5 Regressions including the novel combination characteristics (Standard errors clustered around the researcher)

VARIABLES	(1) log(Citations)	(2) log(Citations)	(3) log(JIF)	(4) log(JIF)
Novel article	-0.31	-0.28	-0.37***	-0.27**
Follower article 1-2yrs	0.091	0.087	-0.22***	-0.16***
Follower article 3-4yrs	-0.025	-0.0032	-0.22***	-0.17***
Novel article*Universe n. authors	0.068**	0.066**	0.024	0.024
Novel article*Universe n. articles	0.031	0.017	0.045	0.043
Novel article*Universe US author	0.021	-0.014	-0.037	-0.099
Article n. of authors	0.011***	0.0098***	-0.00065	0.00017
Article US affiliation	0.21***	0.22***	0.091***	0.088***
N. of references (log)	0.54***	0.55***	0.24***	0.23***
Female scientist	-0.040		-0.021	
CNRS scientist	0.048		0.028	
Scientist Age	-0.0065***	0.084***	-0.0034***	0.048**
Scientist's publication stock (log)	0.21***	-0.47***	0.29***	-0.17
Scientist's average citations (log)	0.24***	-0.40***	-0.021	0.11***
Scientist's average IF (log)	-0.028	0.095	0.39***	-0.55***
Condensed matter	0.047		0.062***	
Constant	-0.47***	-1.82***	0.12	-0.43
Observations	11,753	11,753	11,753	11,753
R-squared	0.174	0.129	0.122	0.052
Dummy year	yes	yes	yes	yes
Scientist fixed effects	no	yes	no	yes

Table 6 Regressions on citations at t=0,1,2,3,4 including the novel combination characteristics (Standard errors clustered around the researcher)

VARIABLES	(1) log(Cit t=0)	(2) log(Cit t=0)	(3) log(Cit t=1)	(4) log(Cit t=1)	(5) log(Cit t=2)	(6) log(Cit t=2)	(7) log(Cit t=3)	(8) log(Cit t=3)	(9) log(Cit t=4)	(10) log(Cit t=4)
Novel article	-0.00095	-0.16	-0.051	-0.18	0.036	-0.29	0.16**	0.065	0.14**	-0.31
Follower article 1-2yrs	0.049	0.048	0.048	0.048	0.067	0.067	0.11**	0.11**	0.17***	0.17***
Follower article 3-4yrs	0.048	0.048	-0.084*	-0.084*	0.055	0.055	0.025	0.025	0.10**	0.10**
Novel article*Universe n. authors		0.021		0.013		0.063***		0.0049		0.053**
Novel article*Universe n. articles		0.053		0.021		-0.0097		0.026		0.078
Novel article*Universe US author		-0.20**		0.032		0.14		-0.004		-0.0059
Article n. of authors	0.0049***	0.0048***	0.0079***	0.0079***	0.0091***	0.0091***	0.0086***	0.0086***	0.0083***	0.0082***
Article US affiliation	0.11***	0.11***	0.15***	0.15***	0.15***	0.14***	0.19***	0.19***	0.20***	0.20***
N. of references (log)	0.13***	0.13***	0.36***	0.36***	0.37***	0.37***	0.37***	0.37***	0.35***	0.35***
Scientist Age	0.045***	0.044***	0.077***	0.077***	0.049***	0.049***	0.035*	0.035*	0.049***	0.049***
Scientist's publication stock (log)	-0.16***	-0.16***	-0.39***	-0.39***	-0.31***	-0.31***	-0.25**	-0.25**	-0.30***	-0.30***
Scientist's average citations (log)	-0.044	-0.045	-0.24***	-0.24***	-0.32***	-0.32***	-0.34***	-0.34***	-0.39***	-0.39***
Scientist's average IF (log)	0.079*	0.080*	0.017	0.017	0.15**	0.16**	0.095	0.095	0.12	0.12
Constant	-1.76***	-1.74***	-2.35***	-2.35***	-1.24**	-1.22**	-0.72	-0.72	-1.22**	-1.20**
Observations	11,753	11,753	11,753	11,753	11,753	11,753	11,753	11,753	11,753	11,753
R-squared	0.04	0.041	0.098	0.098	0.097	0.097	0.101	0.101	0.101	0.101
Dummy year	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Scientist fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Conclusion

This paper investigates the impact of publishing an article introducing a novel scientific idea and of publishing follower articles reusing that idea. The analysis is conducted on the articles published by French physicist between 2005 and 2009.

Novel articles receive the same number of citations as non-novel articles in the five years after their publication. The scientific community does not recognize a citation premium to the articles proposing novel scientific ideas. Interestingly, if the novel idea is introduced by multiple authors, there is a significant citation premium. The presence of multiple authors might increase the visibility of the novelty included in the article. When we distinguish short and long run citations, we find that novel and follower articles receive more citations than the non-novel articles in the long run (after 3 / 4 years since the publication).

The quality of the journals where novel and follower articles are published, measured by their impact factor, is lower than the quality of the journals where non-novel articles are published. We can interpret this last finding as a bias of the journals against novel articles.

References

Ayoubi, C., Pezzoni, M., Visentin, F., n.d. At the origins of learning: Absorbing knowledge flows from within the team. *J. Econ. Behav. Organ.*

Fleming, L., 2001. Recombinant Uncertainty in Technological Search. *Manag. Sci*. 47, 117–132.

Merton, Robert K. 1968. "The Matthew Effect in Science." Science 159 (3810): 56–63.

Schumpeter, J. A. (1939). Business cycles (Vol. 1). McGraw-Hill New York.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical Combinations and Scientific Impact. *Science* 342, 468–472.

Verhoeven, D., Bakker, J., Veugelers, R., 2016. Measuring technological novelty with patent-based indicators. *Res. Policy* 45, 707–723.

Wang, J., Veugelers, R., Stephan, P., 2017. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Res. Policy* 46, 1416–1436.