Journal of Informetrics 18 (2024) 101564

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Comparing patent in-text and front-page references to science

Jian Wang [a,b,c,*], Suzan Verberne [b]

[a] Lancaster University Leipzig, Nikolaistrasse 10, 04109 Leipzig, Germany
[b] Leiden Institute of Advanced Computer Science, Leiden University, Einsteinweg 55, Leiden, 2333 CA, Netherlands
[c] Centre for Science and Technology Studies, Leiden University, Kolffpad 1, 2333 BN Leiden, Netherlands

ARTICLE INFO

Keywords:
Science technology linkage
Patent reference
Basicness
Interdisciplinarity
Novelty
Scientific citation

ABSTRACT

Patent references to science provide a paper trail of knowledge flow from science to innovation, and have attracted a lot of attention in recent years. However, we understand little about the differences between two types of patents references: front-page vs. in-text. While both types of references are becoming more accessible, we still lack a systematic understanding on how results are sensitive to which type of references are being analyzed in science and innovation studies. Using a dataset of 33,337 USPTO biotech utility patents, their 860,879 in-text and 637,570 front-page references to Web of Science journal articles, we found a remarkable low overlap between these two types of references. We also found that in-text references are more basic and have more scientific citations than front-page references. The difference in interdisciplinarity and novelty is small when comparing at the reference level and insignificant when comparing at the patent level. We analyze the association between patent value (as measured by patent citations and market value) and characteristics of referenced sciences. Results are substantially different between in-text and front-page references. In addition, in-text referenced papers have a higher chance of being listed on the front-page of the same patent when they are moderately basic, less interdisciplinary, less novel, and have more scientific citations.

1. Introduction

References in patents to scientific papers (often referred to as scientific Non-Patent References, sNPRs) provide a paper trail of the knowledge flow from science to technological innovation. Since the pioneer work of Nunn and Oppenheim (1980), Narin and Noma (1985), and Tamada, Naito, Kodama, Gemba, and Suzuki (2006), sNPRs have been widely used for studying the interaction between science and technology (Ahmadpoor & Jones, 2017; Callaert et al., 2014; Hicks, Breitzman, Hamilton, & Narin, 2000; Ke, 2020, 2023; Poege, Harhoff, Gaessler, & Baruffaldi, 2019; Popp, 2017; van Raan, 2017; van Raan & Winnink, 2018; Veugelers & Wang, 2019).

Until recently, however, most studies have used patent *front-page references* and ignored patent *in-text references. Front-page references* are the references listed on the front page of the patent document, which are deemed as relevant prior art for assessing patentability. *In-text references* are references embedded in patent full text, serving a similar role as references in scientific papers. This focus on front-page references is probably due to data availability. Front-page references are easily retrievable from the meta-data of patents, although matching them to individual scientific papers is not an easy task. In comparison, extracting patent in-text references is a much more challenging task, as these references are embedded in the running text without consistent structural cues and typically contain even less information than front-page references. Thanks to the recent advancement in computer science, scholars have started

* Corresponding author.
E-mail address: jian.wang@lancaster.ac.uk (J. Wang).

https://doi.org/10.1016/j.joi.2024.101564
Received 30 January 2024; Received in revised form 12 June 2024; Accepted 16 July 2024
Available online 19 July 2024

developing rule-based and machine-learning methods for extracting and matching patent in-text references to scientific papers (Bryan, Ozcan, & Sampat, 2020; Marx & Fuegi, 2022; Verberne, Chios, & Wang, 2019; Voskuil & Verberne, 2021).

These recent studies have shown that in-text references embody information that is different from front-page references. All these studies have observed a low overlap between front-page and in-text references, reflecting their distinct generation processes. Front-page references are generated because of patent applicants' legal duty to disclose prior art that is relevant for assessing the patentability of the focal invention, as well as examiners adding such relevant prior art. In comparison, in-text references are more like references in academic papers and may capture prior research that has enabled or influenced the focal invention but does not directly relate to patentability, for example, motivating facts, open scientific puzzles, and enabling research tools (Bryan et al., 2020; Meyer, 2000). Consistent with the process through which references are generated, several earlier studies based on surveys or interviews have suggested that front-page references may not represent a direct link between the citing patent and the cited scientific paper, but the cited scientific paper plays a more indirect role as a source of relevant background information (Callaert, Pellens, & Van Looy, 2014; Meyer, 2000; Nagaoka & Yamauchi, 2015; Tijssen, Buter, & van Leeuwen, 2000). In their work that pioneered the analysis of patent front-page references, Narin and Noma (1985) stated that patent in-text references might be a better instrument for tracing knowledge flow from science to technology. Bryan et al. (2020) provided further evidence supporting this statement: First, using paper-patent-pairs (which consists of a scientific paper and a patent about the same biotech research output), references in the paired paper are much more likely to be cited by the paired patent in text rather than on front page. Second, using firm survey data, the number of a company's patent in-text references to science is more strongly correlated with its reliance on science as reported by its R&D manager. In addition, they also examined how the number of in-text and front-page references is associated with patent value, in terms of patent citations, market value, being in a triad family, and being maintained to year 8.

In addition, Marx and Fuegi (2022) found that in-text references are older, less localized, less self-cited, more interdisciplinary, and more cited by future patents, compared with front-page references. Verluise, Cristelli, Higham, and de Rassenfosse (2020) compared patent in-text references to patents with front-page references to patents. They found a low overlap between them too and that in-text referenced patents have higher text-similarity to the citing patents than front-page referenced patents. Although this study is about patent references to patents rather than to scientific papers, it is still informative for this study as it suggests that in-text referencing exhibits a stronger intellectual connection than front-page referencing, in general.

More recently, large-scale datasets of in-text references are also becoming available (Bryan et al., 2020; Marx & Fuegi, 2022), and researchers have the freedom to choose which type of reference to use or combine them together. Although prior studies have documented differences between front-page and in-text references, we still lack a systematic understanding of how results or findings might be contingent on the type of references being analyzed. This is the central question that this paper aims to answer, and answers to this question have important implications for studies in various areas using patent references data.

More specifically, we focus on four characteristics of science: basicness, interdisciplinarity, novelty, and scientific citations. We focus on basicness and interdisciplinarity because of their significance in science policy discussions regarding how to make science more useful for the society, namely, a key thesis in Gibbons (1994)'s proposal for making science more useful for the society is to move from basic research to applied research, and from discipline-based mode of knowledge production to a transdisciplinary one. We focus on novelty and scientific citations because of their fundamental role in the science reward system. Science, as a self-governed system of work organization and control, is structured to encourage (1) novel contributions to the common stock of scientific knowledge, and (2) research that can influence, direct, and is useful for the work of colleagues in the field, which is reflected in peer recognition and partly scientific citations (Merton, 1973; Stephan, 1996; Whitley, 2000).

Furthermore, prior studies have examined whether basic, interdisciplinary, and novel scientific papers are more likely to be cited by patents (Hicks et al., 2000; Ke, 2020, 2023; Popp, 2017; Veugelers & Wang, 2019). Less explored is how building on basic, interdisciplinary, or novel scientific papers would impact the value of the produced patent. In addition, prior studies have consistently documented that scientific papers that are highly cited in science are also more likely to be cited by patents, but mixed results are observed regarding whether citing highly cited science leads to highly cited patents (Gittelman & Kogut, 2003; Poege et al., 2019). We contribute to this stream of literature by exploring the following research questions: (1) are in-text references more (or less) basic, interdisciplinary, novel, and highly cited than front-page reference? (2) do in-text and front-page references lead to the same conclusions regarding the association between patent value (as measured by patent citations and market value) and characteristics of references science (in terms of basicness, interdisciplinarity, novelty, and scientific citations), and (3) what types of in-text references are more likely to be listed on the front-page of the same patent?

In short, this paper aims to make two contributions: First, it informs future science and innovation studies about the use of patent references, by systematically assessing the differences between front-page and in-text references and especially whether they lead to different analytical results. Second, it explores the association between the characteristics of cited science (in terms of basicness, interdisciplinarity, novelty, and scientific citations) and patent value, beyond the question that has been examined in the prior literature regarding which types of science is more likely to be cited by patent.

2. Data and method

2.1. Data

Our paper-patent-links data come from Voskuil and Verberne (2021). They trained the state-of-the-art BERT-based machine-learning models for extracting patent in-text references to scientific papers. The pre-trained BERT models were fine-tuned on a set of 1,952 hand-labeled references in 22 patent documents. The algorithm automatically classified words into three categories: (*B*) the beginning of a reference, (*I*) inside a reference, and (*O*) outside a reference. The accuracy of the algorithm is reflected in its prediction power for *B* and *I* labels (Ramshaw & Marcus, 1999; Sang & De Meulder, 2003). The accuracy of the best-performing method, as measured in leave-one-out validation, is very high: test recall and precision are 94.7 % and 95.4 % respectively for beginnings of citations, and 98.6 % and 97.6 % for words inside citations. Subsequently, they matched the extracted in-text references (as well as front-page references) to journal articles in Web of Science (WoS) using rule-based reference parsing. The final dataset consists of all the biotech utility patents granted by USPTO from 2006 to 2010 retrieved from Google Patents, and each patent is linked to a set of its referenced WoS journal articles, in text or on front-page. In total, the dataset for our analysis consists of 33,337 patents and their 860,879 in-text and 637,570 front-page references to scientific papers in WoS.[1] Building on the paper-patent-link data, we first construct various measures for individual patents and papers. For regression analysis, our unit of analysis is a patent. We use two measures of patent value as dependent variables and construct science measures based on the profile of a patent's (in-text or front-page) scientific references as independent variables.

2.2. Science measures

We construct measures for individual scientific papers. For *Basicness*, we adopt the measure proposed by Weber (2013) for biomedical research, which classifies a paper as *highly basic* if it only has cell/animal-related MeSH terms but no human-related MeSH terms, *moderately basic* if it has both cell/animal- and human-related MeSH terms, and *not basic* (i.e., clinical) if it only has human-related but not cell/animal-related MeSH terms. This measure is an ordinal measure, and its attributes 1, 2, and 3 correspond to *not basic, moderately basic*, and *highly basic*, respectively. This measure is specific for biomedical fields and is suitable for our sampled patents in the domain of biotech. Ke (2020) used this measure and found that papers that are more basic have a higher chance of being cited by patents. We use the PubMed ID provided by WoS to extract papers in the PubMed database and their MeSH terms.

For *Interdisciplinarity*, we adopt the Rao-Stirling measure (Stirling, 2007; Wang, Thijs, & Glänzel, 2015), which captures all the three diversity dimensions (i.e., variety, balance, and disparity) of the involved disciplines underlying a study. More specifically, it equals $\sum_{i \neq j} p_i p_j d_{ij}$, where i and j are indices of a paper's referenced disciplines (i.e., WoS subject categories), p_i is the proportion of references to discipline i, and d_{ij} is the distance between discipline i and j, measured as 1 – cosine similarity between discipline i and j based on their co-citation matrix. This measure is a continuous measure ranging from 0 to 1. Ke (2023) used this measure and found that papers that are more interdisciplinary are more likely to be cited by patents.

For *Novelty*, we adopt the measure developed by Wang, Veugelers, and Stephan (2017), which follows the combinatorial novelty perspective and identifies novel paper as the ones that makes unprecedented combinations of pre-existing knowledge components, where knowledge components are proxied by referenced journals. This measure is a binary variable: 1 if novel and 0 if not novel. Veugelers and Wang (2019) used this measure and find that papers that are more novel are more likely to be cited by patents.

For *Scientific citations* we count the number of forward citations a scientific paper receives from future papers in the Web of Science (WoS) database, using a five-year citation time window, following the common practice (Wang, 2013). Citation counts have been widely used as the focal explanatory variable or a control variable in previous sNPR studies. While prior studies have provided consistent evidence that highly cited scientific papers are more likely to be cited by patents (Ahmadpoor & Jones, 2017; Hicks et al., 2000; Popp, 2017; Veugelers & Wang, 2019), findings are mixed regarding the association between impactful scientific papers and impactful patents (Gittelman & Kogut, 2003; Poege et al., 2019).

2.3. Patent measures

For each patent, we construct two measures to capture its value: (1) *Patent citations*, which is the number of times a patent is cited by future patents, using a five-year citations time window, following the common practice. (2) *Market value*, which is based on the stock market response to the issuing of the patent, in million US dollars, developed by Kogan, Papanikolaou, Seru, and Stoffman (2017). We downloaded the market value data from https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data. Our analysis uses the real value deflated to 1982 (million) dollars using the CPI. Information about the market value is available for a subset of 8,392 patents in our sample.

When investigating the association between patent citations and the characteristics of its cited science, in terms of basicness, interdisciplinarity, novelty, and scientific citations, we take the average of these four measures across its referenced scientific papers: *Avg*(*Basicness*), *Avg(Interdisciplinarity*), *Avg*(*Novelty*), and *Avg*(*Scientific citations*). In addition, our focal explanatory variables also include *I*(*Having references*), which indicates whether a patent has any scientific references, and *# References*, which is the number of unique WoS papers referenced by a patent. For all these measures, we construct two versions: one based on patent in-text references and the other based on patent front-page references.

[1] The dataset from Voskuil and Verberne (2021) consists of 33,338 patents, we excluded one plant patent and kept only utility patents.

3. Results

3.1. Comparing in-text and front-page references

We first examine the overlap between patent front-page and in-text references. The 33,337 USPTO biotech patents in our sample made 1,325,168 references to WoS papers either in text or on front page. In other words, pooling together in-text and front-page reference uncovers 1,325,168 paper-patent-links. Among them 860,879 are in-text references, and 637,570 are front-page references. Fig. 1 reports the overlap between in-text and front-page references. In total, 173,281 references appear both in the text and on the front page of the same patent, which account for only 20 % of all in-text references and 27 % of all front-page references. This observed low overlap is in line with prior observations and suggests that in-text and front-page references embody different types of information (Bryan et al., 2020; Marx & Fuegi, 2020; Verberne et al., 2019).

A scientific paper can be cited by multiple patents. The 1,325,168 total references are linked to 336,522 unique papers, the 860,879 in-text references are linked to 195,988 unique papers, and the 637,570 front-page references are linked to 245,852 unique papers. Although in-text references have a larger volume (i.e., more paper-patent-links), they are linked to fewer unique papers, compared with front-page references. In other words, in-text references are concentrated in a smaller set of papers than front-page ones. In-text referenced papers are cited more often than front-page referenced papers. On average, in-text referenced papers are cited by 5.4 patents in our sample in text or on front page, 4.4 patents in text, and 1.9 patents on front-page, and the corresponding numbers are 4.6, 2.7 and 2.6 for front-page referenced papers, respectively.

We further assess the difference between in-text and front-page references in terms of their basicness, interdisciplinarity, novelty, and scientific citations. Fig. 2 plots the distributions of these four measures for in-text and front-page references separately. Because the sample size is large, all the mean differences are highly significant (i.e., $p < 0.001$) according to Welch two sample t-tests and Wilcoxon rank sum tests, although the difference in interdisciplinarity and novelty seem very small in size. Taken together, results show that in-text references are more basic and have more scientific citations than front-page references. In-text references are less interdisciplinary but more novel than front-page references, but the differences are small.

3.2. Patent level comparison

Descriptive statistics for patent-level variables are reported in Table 1. 80.6 % of our sampled patents have in-text scientific references, while 87.3 % have front-page references. Among those with in-text references, they cite on average 32.0 scientific papers in-text. Among those with front-page references, they cite on average 21.9 papers on front-page. These differences are significant according to Wilcoxon matched-pairs signed-rank test ($p < 0.001$, Table 2). In the previous section, we have shown that in-text references have a larger volume but are concentrated in a smaller set of scientific papers. It appears that in-text references are also concentrated in a smaller set of patents.

Wilcoxon matched-pairs signed-rank tests (Table 2) also suggest that the average basicness and scientific citations of papers in a patent's in-text references are significantly higher than that of front-page references in the same patent ($p < 0.001$), while there are no significant differences in average interdisciplinarity ($p = 0.889$) or novelty ($p = 0.082$).

Correlations between the variables based on in-text and front-page references are moderate (Table 2). The Spearman correlation is 0.466 between whether having in-text references and whether having front-page references. The correlations between two versions of variables (i.e., in-text and front-page) are 0.324, 0.602, 0.533, 0.261, and 0.430, for the number of referenced papers, average basicness, average interdisciplinarity, average novelty, and average scientific citations, respectively. These moderate correlations suggest that, if we rank patents by their number of scientific references or the average basicness, interdisciplinarity, novelty, and scientific citations of their referenced scientific papers, using in-text and front-page references will produce rankings that are substantially different. Furthermore, if we study the association between the characteristics of patents and the characteristics of their

Fig. 1. Overlap between in-text and front-page references.

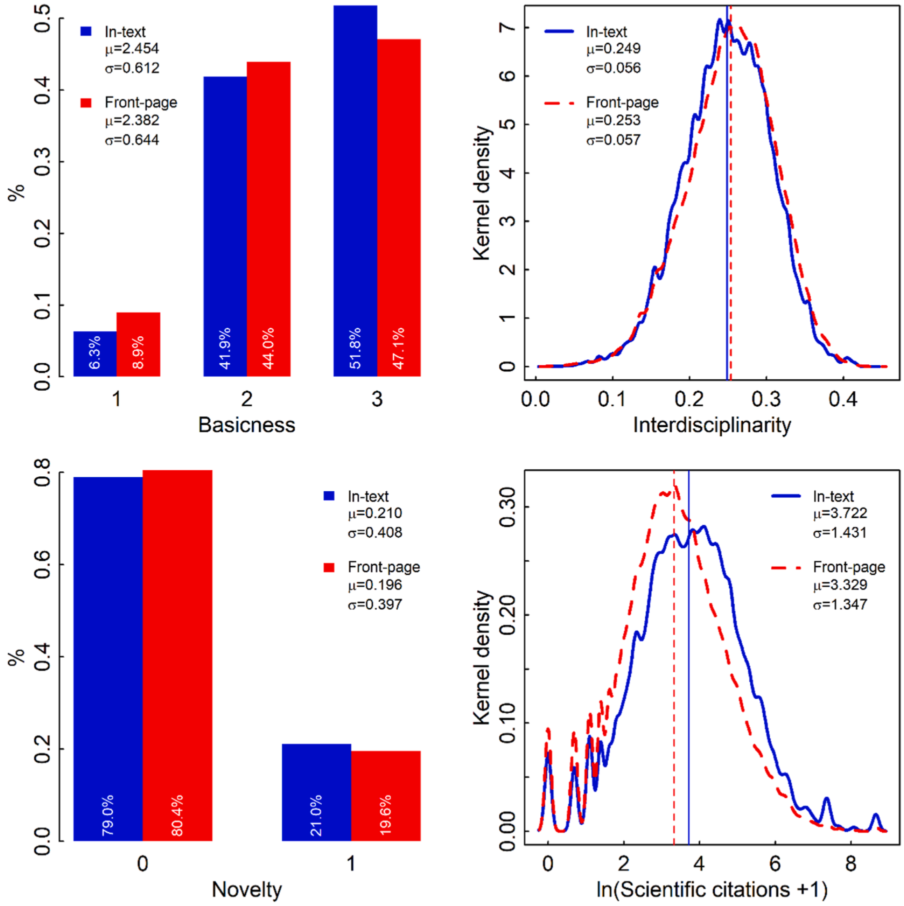

Fig. 2. Distribution of basicness, interdisciplinarity, novelty, and scientific citations, by in-text and front-page references. Plots for basicness and novelty are simple proportions by category. Plots for interdisciplinarity and *ln*(Scientific citations +1) are kernel densities where the vertical lines mark the mean values. Scientific citations are logarithm transformed because it is highly skewed.

Table 1
Descriptive statistics (Unit of analysis: patent).

	N	Mean	Std. Dev.	Min	Max
Patent citations	33,337	3.766	8.238	0	549
Market value (m$)	8,392	19.784	35.222	0.000	445.335
In-text					
I(Having references)	33,337	0.806	0.395	0	1
# References	26,872	32.036	43.005	1	711
Avg(Basicness)	26,012	2.445	0.394	1	3
Avg(Interdisciplinarity)	26,709	0.254	0.032	0.037	0.429
Avg(Novelty)	26,872	0.155	0.157	0	1
Avg(Scientific citations)	26,872	124.208	191.441	0	5871
Front-page					
I(Having references)	33,337	0.873	0.333	0	1
# References	29,110	21.902	31.368	1	1064
Avg(Basicness)	27,999	2.401	0.435	1	3
Avg(Interdisciplinarity)	28,982	0.256	0.037	0.023	0.419
Avg(Novelty)	29,110	0.162	0.178	0	1
Avg(Scientific citations)	29,110	62.580	88.399	0	5871

Table 2
Patent level comparison.

	Wilcoxon matched-pairs signed-rank test		Spearman correlation
	z	p	
I(Having references)	-31.765	0.000	0.466
# References	25.025	0.000	0.324
Avg(Basicness)	17.549	0.000	0.602
Avg(Interdisciplinarity)	0.139	0.889	0.533
Avg(Novelty)	-1.739	0.082	0.261
Avg(Scientific citations)	83.702	0.000	0.430

referenced scientific papers, we might come to different conclusions depending on whether in-text or front-page references are used.

3.3. In-text scientific references and patent value

In the next step we investigate the association between the characteristics of in-text referenced science and patent citations. The dependent variable is an over-dispersed count variable, so we fit a series of Negative Binomial (NB) models. Regression results are reported in Table 3. Column 1 reports the NB model that uses whether having scientific references as the focal independent variable and incorporates the complete set of patent's issuing year and IPC subclass (4-digit level) dummies. The result suggests that patents having in-text scientific references receive 29.1 % more patent citations than patents not having in-text scientific references, issued in the same year and IPC subclass. Within the set of patents that have in-text scientific references, we further examine the intensity of reliance on science, that is, the number of referenced scientific papers. This independent variable is also a count variable and has a skewed distribution, so we take its natural logarithm for regression analysis. Column 2 shows that as a patent's number of referenced papers increases by 1 %, its patent citations increase by 0.122 %.

Then we move on to explore the characteristics of in-text referenced science. Column 3-6 each uses average basicness, interdisciplinarity, novelty, and scientific citations of referenced papers as the focal independent variable. In all these models, the *ln* number of scientific references is controlled for, in addition to patent issuing year and IPC subclass. *Avg*(*Scientific citations*) is skewed so it takes natural logarithm transformation for regression analysis. Column 3 shows that, as the average basicness of referenced papers increases by 1, patent citations decrease by 7.0 %, holding all other variables fixed. Column 4 suggests no significant association between patent citations and interdisciplinarity. Column 5 shows that, as the average novelty of referenced papers increases by 1, patent citations increase by 15.6 %, holding all other variables fixed. Column 6 suggests no significant association between patent citations and scientific citations. Column 7 further fits a model with all these four variables together and yields consistent results as running separate models for each independent variable (i.e., Column 3-6). In summary, patents building on less basic but more novel science are more impactful in the technological domain.

We then use the market value (in million US dollars) of the patent based on stock market reaction to the event of patent being issued as the dependent variable. This variable is also skewed but not a count variable, so we take natural logarithm transformation and then fit Ordinary Least Squares (OLS) models. Results are reported in Table 4. Results show that patents with in-text scientific references worth 80.2 % more than patent without in-text scientific references (Column 1). Within the set of patents having in-text scientific references, patent market value increases by 0.146 % as the number of referenced scientific papers increases by 1 % (Column 2). As the average basicness of referenced science increases by 1, patent market value decreases by 31.6 % (Column 3). Interdisciplinarity and

Table 3
In-text scientific references and patent citations.

	Patent citations NB						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I(Having references)	0.291*** (0.029)						
ln(# References)		0.122*** (0.011)	0.133*** (0.012)	0.122*** (0.011)	0.122*** (0.011)	0.124*** (0.011)	0.136*** (0.012)
Avg(Basicness)			-0.070* (0.034)				-0.073* (0.035)
Avg(Interdisciplinarity)				0.622 (0.446)			0.219 (0.506)
Avg(Novelty)					0.156^{+} (0.082)		0.175* (0.086)
ln(Avg(Scientific citations) +1)						-0.009 (0.013)	-0.009 (0.014)
Issue year	Y	Y	Y	Y	Y	Y	Y
IPC subclass	Y	Y	Y	Y	Y	Y	Y
N	33337	26872	26012	26709	26872	26872	25930
BIC	152066	123120	118841	122524	123115	123130	118555

All science measures are based on in-text references. Robust standard errors in parentheses.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $^{+}p < 0.10$.

Table 4
In-text scientific references and patent market value.

	ln(Market value) OLS						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I(Having references)	0.802*** (0.073)						
ln(# References)		0.146*** (0.018)	0.147*** (0.019)	0.136*** (0.018)	0.146*** (0.018)	0.125*** (0.019)	0.124*** (0.020)
Avg(Basicness)			-0.316*** (0.073)				-0.321*** (0.074)
Avg(Interdisciplinarity)				-0.761 (0.855)			-0.181 (0.999)
Avg(Novelty)					0.090 (0.170)		0.021 (0.200)
ln(Avg(Scientific citations)+1)						0.073** (0.026)	0.085** (0.028)
Issue year	Y	Y	Y	Y	Y	Y	Y
IPC subclass	Y	Y	Y	Y	Y	Y	Y
N	8392	7061	6834	7023	7061	7061	6819
R2	0.074	0.052	0.054	0.051	0.052	0.054	0.055
BIC	35192	28860	27807	28702	28869	28859	27758

All science measures are based on in-text references. Robust standard errors in parentheses.
***p < 0.001; **p < 0.01; *p < 0.05; $^{+}$p < 0.10.

novelty has no significant association with patent market value (Column 4 and 5). As the average scientific citations of referenced science increase by 1 %, patent market value increases by 0.073 % (Column 6). These results are robust when fitting a model with all four variables together (Column 7). In summary, patents building on papers that are less basic but more highly cited in science have higher private market value.

3.4. Complexity in the association between patent value and referenced science

Our regression models assume a linear equation, where the left-hand side is the natural log of one dependent variable (patent citations or market value), and the right-hand side consists of a series of independent variables. This setup is flexible for fitting positive (or negative) effects at an increasing or decreasing rate. However, it does not allow nonmonotonic effects (e.g., inverted U-shaped) or discontinuous effects. Therefore, we reexamine the associations using a non-parametric approach without assuming a linear equation. Specifically, we categorize our independent variables into 10 ordered and evenly sized deciles and then estimate the expected patent

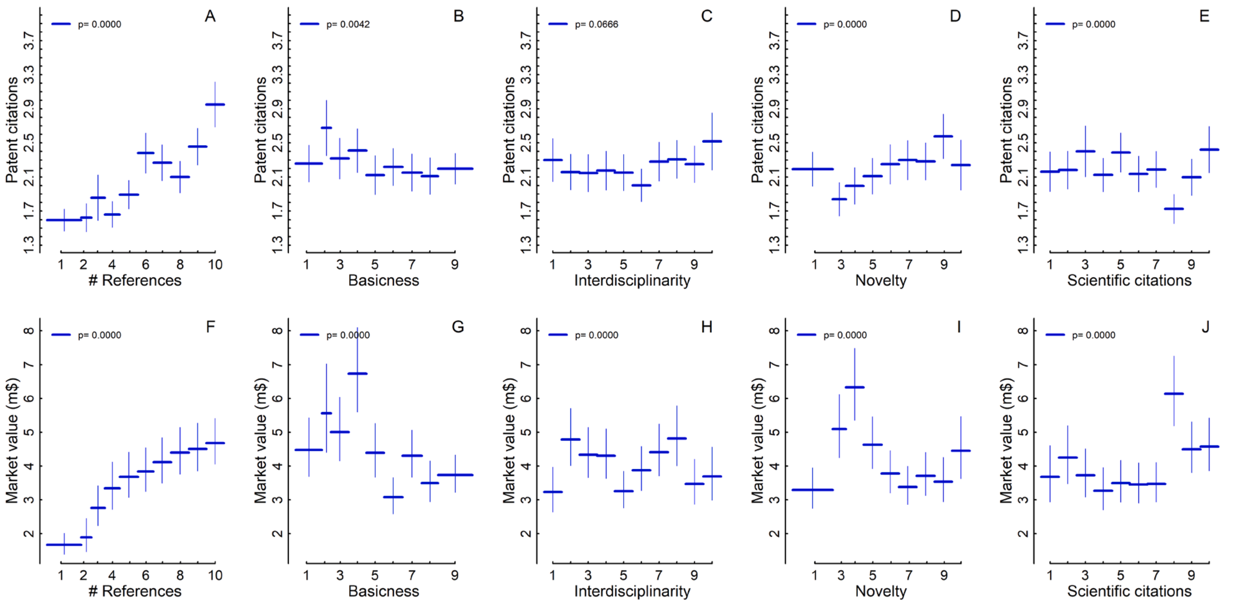

Fig. 3. In-text scientific references and patent value. This figure plots the estimated value of patent value for an average patent in different science measure deciles. An average patent means its issuing year is 2010, IPC is C12N, and for Plot B-E and G-J, the natural log number of references takes the mean value. The length of the blue horizontal lines is proportional to the decile size. The blue vertical lines mark the 95 % confidence interval. p-value is for the joint significance of all the levels of the decile variable.

citations and market value for each decile. Because of ties, not all deciles are evenly sized. Taking the number of in-text scientific references as an example, 6,465 patents with 0 references are classified into Decile 1, the next 1,891 patents with only 1 reference are classified into Decile 2, …, the last 3,330 patents with 71 to 711 references are classified into Decile 10. Then we use decile assignment as a factor/categorical variable for regression analysis, and results are reported in Online supplementary materials Table S2 and S3. These regressions essentially estimate the differences between the reference decile (i.e., Decile 1, omitted in the regression table) and the other nine deciles (Decile 2-9 reported in the regression table), controlling for other control variables. Based on the regression results, we can estimate the expected value of the dependent variables for an average patent in each decile (i.e., issuing year is 2010 (largest in our sample), IPC subclass is C12N (largest in our sample), and other control variables (if any) at the mean). Fig. 3 plots these estimates.

Regarding the number of in-text references, consistent with the result reported in the preceding section, Fig. 3A&F display a roughly continual increase in patent citations and market value, as the number of in-text references move from 0 (Decile 1) to 1 (Decile 2), and then further increases (from Decile 2 to Decile 10). There is consistent evidence that citing science and citing more scientific papers is associated with higher patent value.

While the linear model reported in the preceding section shows a negative association between basicness and patent value, results in Fig. 3B&G suggest an inverted U-shaped pattern. Both patent citations and market value first increase and then decrease as the average basicness increases. Patent citations reach the peak point at Decile 2, and if we dismiss Decile 2 due to its small decile size, then the peak point is reached at Decile 4. Patent market value reaches its peak at Decile 4. These results suggest that a moderate level of basicness is associated with the highest patent value.

While the linear model suggests an insignificant association between interdisciplinarity and patent value, Fig 3C suggests a U-shaped effect, but the p-value for the joint significance of interdisciplinarity deciles is larger than 0.05. Fig 3H suggests no clear association between average interdisciplinarity and patent market value. Taken together, we conclude no consistent or significant association between interdisciplinarity and patent value.

The linear model suggests that novelty is positively associated with patent citations but insignificantly with patent market value. Fig. 3D&I reveal more complex patterns. According to Fig 3D, as a patent moves from having no novel references to having novel references, there is a sudden drop in patent citations. As the average novelty further increases, patent citations rise and slowly reach a plateau (or even go down). According to Fig 3I, there is a disruptive rise in patent market value when moving from having no novel references to having novel references. However, as the average novelty further increases, patent market value decreases and slowly flattens (or even bounces up). Taken together, these results suggest a structural change between patents building on novel science and those not. As a patent builds on novel science, its technological impact drops, potentially due to uncertainties introduced by sourcing novel science. Its technological impact then recovers and reaches a higher point than patents not building on novel science, indicating that sourcing more novel science leads to broader and more unexpected applications. However, this increasing trend does not continue unlimitedly, the benefit from sourcing novel science stops at certain level of average novelty. Regarding patent market value, sourcing novel science brings a jump in the stock market reaction to the patented technology, reflecting market's appreciation of novelty. However, further increase in the novelty of sourced science reduces market value as the patent might become too remote from

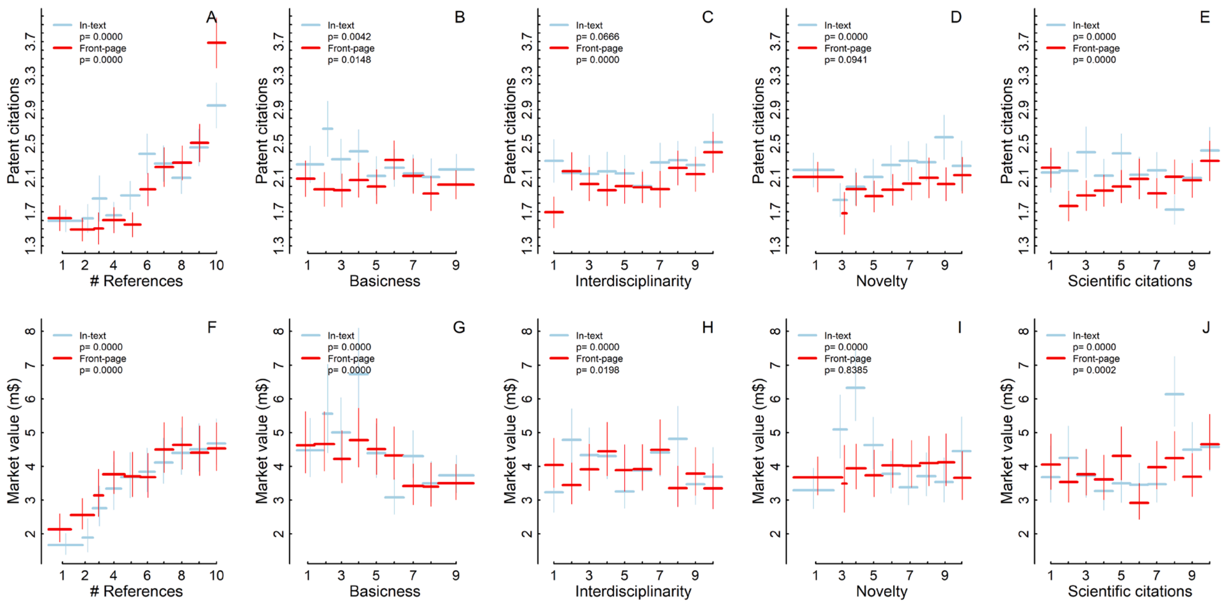

Fig. 4. Scientific references and patent value: Front-page vs. in-text. The light blue lines in this plot are identical to the blue lines in Fig. 3. They represent results based on in-text references. We overlay them with the red lines (results based on front-page references) for an easier comparison. Specifically, we repeat the same procedure for producing Fig. 3 but use science measures based on patent front-page references instead.

marketable applications.

Consistent with results based on the linear model, Fig 3E exhibits no clear associations between scientific citations and patent citations, but only fluctuations around a flat line, and Fig 3J displays an increasing trend with some fluctuations. Therefore, we conclude that scientific citations have no significant association with patent citations but a positive association with patent market value. This suggests that the quality standard or taste might not be perfectly aligned between science and technology, scientific outputs that are (perceived) useful for others to do follow-on scientific research (i.e., receive more scientific citations) do not necessarily leads to technologies that are (perceived) useful for others to develop follow-on technologies (i.e., receive more patent citations). On the other hand, there is neither a conflict between them as no significantly negative relation is observed. Regarding patent market value, scientific outputs that are highly recognized by other scientists are positively associated with technologies that are highly appreciated by the stock market, reflecting a certain level of alignment between scientists' and the market's quality standard or taste.

3.5. *Front-page scientific references and patent value*

We repeat all the analyses using front-page references instead, to test whether using front-page will lead to the same findings. Fig. 4 repeats Fig. 3 but additionally overlays it with estimates based on patent front-page references. Tables 5 and 6 replicate regression analysis reported in Tables 3 and 4, respectively.

Results on the number of front-page scientific references are consistent with what is observed when using in-text scientific references (Fig. 4A&F), which display a continually positive association between the number of scientific references and patent value. Regarding average basicness, results are substantially different. While the inverted U-shaped association between basicness and patent market value with an overall negative trend remains (Fig 4G), there are no clear associations between average basicness and patent citations (Fig 4B). While the insignificant association between interdisciplinarity and patent market value remains (Fig 4H), there is a significantly positive association between interdisciplinarity and patent citations (Fig 4C). No clear patterns or significant associations between average novelty and patent value are displayed (Fig 4D&I). Fig4D further shows that the general trend based on front-page references still resembles that of in-text references, but the size of association is much smaller and becomes insignificant. While the positive association between scientific citations and patent market value remains (Fig 4J), the association between scientific citations and patent citations becomes weakly significantly positive (Fig 4E).

In summary, using front-page reference cannot replicate all the results based on in-text references. More specifically, when examining patent market value, results based on front-page references are largely consistent with results based on in-text references, except that the association with novelty becomes insignificant. When examining patent citations, results are substantially different: Association with basicness and novelty become insignificant, while insignificant associations with interdisciplinarity and scientific citations become significantly positive.

3.6. *Why do front-page and in-text references yield different results?*

The inconsistencies between the results based on front-page and in-text references are not surprising, considering their low overlap and the moderate correlations between science measures based on front-page and in-text references. In this section we attempt to explore why such inconsistencies emerge, by looking into the processes through which in-text and front-page references are generated. As discussed before, in-text references document various sources of knowledge that are instrumental to the patented technology, while front-page references are listed for disclosing prior arts that are relevant for assessing patentability. Sampat (2010) argued that, for

Table 5
Front-page scientific references and patent citations.

	Patent citations NB						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I(Having references)	0.229*** (0.035)						
ln(# References)		0.210*** (0.010)	0.222*** (0.010)	0.213*** (0.010)	0.210*** (0.010)	0.203*** (0.011)	0.214*** (0.011)
Avg(Basicness)			-0.007 (0.031)				0.029 (0.030)
Avg(Interdisciplinarity)				1.759*** (0.356)			2.038*** (0.393)
Avg(Novelty)					0.102 (0.071)		-0.019 (0.077)
ln(Avg(Scientific citations)+1)						0.025^{+} (0.014)	0.051** (0.015)
Issue year	Y	Y	Y	Y	Y	Y	Y
IPC subclass	Y	Y	Y	Y	Y	Y	Y
N	33337	29110	27999	28982	29110	29110	27959
BIC	152144	132262	127270	131696	132269	132267	127043

This table repeats Table 3 (based on in-text references) but uses science measures based on front-page references instead. Robust standard errors in parentheses.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $^{+}p < 0.10$.

Table 6
Front-page scientific references and patent market value.

	ln(Market value) OLS						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I(Having references)	0.581*** (0.083)						
ln(# References)		0.164*** (0.019)	0.149*** (0.020)	0.163*** (0.019)	0.164*** (0.019)	0.139*** (0.021)	0.128*** (0.021)
Avg(Basicness)			-0.321*** (0.062)				-0.329*** (0.062)
Avg(Interdisciplinarity)				-1.175 (0.713)			-0.508 (0.788)
Avg(Novelty)					0.008 (0.147)		-0.131 (0.159)
ln(Avg(Scientific citations)+1)						0.078** (0.027)	0.067* (0.029)
Issue year	Y	Y	Y	Y	Y	Y	Y
IPC subclass	Y	Y	Y	Y	Y	Y	Y
N	7336	6435	6203	6408	6435	6435	6191
R2	0.078	0.071	0.072	0.071	0.071	0.073	0.073
BIC	31045	26862	25729	26739	26853	26844	25699

This table repeats Table 4 (based on in-text references) but uses science measures based on front-page references instead. Robust standard errors in parentheses.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $^{+}p < 0.10$.

patents that are expected to be more valuable, patent applicants may perform a more comprehensive prior art search, to prevent the chance that the patent application is rejected due to failure of disclosure. This more comprehensive search may result in a longer list of front-page references. Therefore, we view in-text references as an unbiased (but noisy) representation of the scientific outputs underlying a focal patented technology. In comparison, front-page references also reflect this unbiased representation, but they undergo additional screening for relevance to patentaiblity and are subject to biases introduced by patent applicants' strategic behavior. It is possible that certain types of scientific papers are valuable for inspiring the patented technology but not so relevant for assessing its patentability and therefore are not listed on the front page. Applicants might also consciously or unconsciously omit or overreport certain types of references due to strategic considerations or cognitive biases. To explore this, we analyze individual in-text references and examine which types of in-text referenced papers are more likely to be listed on the front page of the same patent.

Using in-text references (i.e., paper-patent-links) as the unit of analysis, we fit conditional fixed-effects logistic models, with patent fixed effects to account for patent heterogeneities. Regression results are reported in Table 7. Column 1 shows that, for the same patent, among its in-text referenced papers, moderately-basic papers have the highest chance of being listed on its front page, followed by highly-basic, and lastly not-basic papers. Columns 2 and 3 show that more interdisciplinary and novel papers among the in-text referenced papers of the same patent are less likely to be listed on the front page of that patent. In contrast, papers with more scientific citations have a higher chance of being listed on the front page of the same patent (Column 4). Results are consistent when all the four focal variables are included in the regression analysis (Column 5).

The inverted U-shaped relationship between basicness and the likelihood of being listed on patent front page is in line with the observed inverted U-shaped relationship between average basicness and patent value. This means that a moderate level of basicness is not only positively associated with higher patent value but also a higher degree of relevance for assessing patentability. Both interdisciplinarity and novelty deviate from the existing paradigm, and their contribution to the patented technology might be rather unexpected. Therefore, their intellectual link to the patent is relatively distant and tenuous, and their relevance for assessing patentability is relatively low. On the other hand, highly cited papers have generated more follow-on research and therefore is also possible to have more direct relevance for assessing patentability (i.e., relevance mechanism). In addition, highly cited papers are more visible in both domains of science and technology, such that missing them would bring a higher risk of being rejected due to failure of disclosure (i.e., visibility mechanism). To further disentangle the relevance vs. visibility mechanism, we incorporate another variable into the analysis, the Impact Factor of the journal where the paper is published. We expect that higher impact factor also increases visibility but not necessarily the relevance for assessing patentability: First, examiners might be more aware of higher impact factor journals, but higher impact factor journals tend to be less applied and accordingly less relevant for assessing patentability (Hamilton, 2003; Noma, 1986), second, while all papers in a high impact factor journal benefit from the visibility of the journal, not all papers will be equally relevant for assessing patentability. We find that papers in higher impact factor journals are indeed much more likely to be listed on the front-page on the same patent (Columns 6-10). More importantly, adding the journal impact factor drastically reduces the size of the coefficient on scientific citations (Column 9) and even renders it insignificant when all four focal science measures are included (Column 10). This suggests that the visibility mechanisms might be driving the observed positive association between scientific citations and the likelihood of being listed on the front page.

Overall, since front-page references systematically under-represent interdisciplinary and novel papers but over-represent moderately basic and highly cited papers, we can expect that using front-page references will yield substantially different results than using in-text references when analyzing these science measures.

Table 7
What types of in-text referenced papers are more likely to be listed on the patent front page?

	I(Front-page) Conditional fixed-effects logit									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Basicness = 1	-0.225***				-0.181***	-0.175***				-0.135***
	(0.019)				(0.020)	(0.020)				(0.020)
Basicness = 2	0.106***				0.099***	0.108***				0.109***
	(0.009)				(0.009)	(0.009)				(0.009)
Interdisciplinarity		-0.638***			-0.579***		-0.349***			-0.341***
		(0.077)			(0.088)		(0.079)			(0.089)
Novelty			-0.025*		-0.022^{+} (0.012)			-0.033**		-0.024* (0.012)
			(0.011)					(0.011)		
ln(Scientific citations +1)				0.065***	0.047***				0.010**	-0.003 (0.004)
				(0.003)	(0.004)				(0.004)	
ln(Journal impact factor)						0.227***	0.249***	0.250***	0.256***	0.224***
						(0.008)	(0.007)	(0.007)	(0.008)	(0.009)
Paper: Publication year	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Paper: Scientific field	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N obs	570408	634225	613065	655311	540379	542245	609058	599218	622234	528070
N patents	15374	16217	16093	16394	15185	15223	16088	16015	16228	15102

Unit of analysis: in-text references, i.e., paper-patent-links through in-text referencing. All models incorporate patent fixed effects, so that estimates are about within-patent differences. The dependent variable I(Front-page) is a binary variable: 1 if an in-text referenced paper is listed on the front page of the same patent, and 0 otherwise. Robust standard errors in parentheses.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $^{+}p < 0.10$.

3.7. Additional analyses

Novelty. We use the novelty measure developed by Wang et al. (2017), who acknowledged the association between novelty and interdisciplinary but argued that novelty is a rarer activity in science than interdisciplinary research. Fontana, Iori, Montobbio, and Sinatra (2020) further warned about the overlap between novelty and interdisciplinarity. Our analysis does show that both novel and interdisciplinary papers are less likely to be listed on the front-page (Section 3.6), even though their effects remain significant when both variables are included in the regression analysis. Furthermore, results are distinct regarding how they are associated with patent value (Sections 3.3, 3.4, 3.5), suggesting that they are not the same thing. Alternative measures of novelty based on MESH terms (Boudreau, Guinan, Lakhani, & Riedl, 2016) or the text of paper title and abstract (Arts, Melluso, & Veugelers, 2023) are available. We also try these measures, more specifically, whether a paper has any new MESH terms, new pairs of MESH terms, new words, or new pairs of words, as well as the share of these new items. Results based on three sources of information (references, MESH, and title/abstract text) are not consistent with each other (Online supplementary materials (SM) §3). Assessing the validity of and differences between various novelty measures itself is an active research topic (e.g., Iori, Pelletier, Souza, Fontana, & Geuna, 2024) and beyond the scope of this paper. Our interpretation of results concerning novelty is based on the assumption that the measure developed by Wang et al. (2017) has satisfactory validity (especially face and construct validity).

Additional controls. Our main results are as follows: using in-text references suggests (1) a positive effect of the number of referenced scientific papers on patent value, (2) an inverted U-shaped and generally negative effect of basicness, (3) an insignificant effect of interdisciplinarity, (4) a discontinuous and nonlinear effect of novelty, and (5) a positive effect of scientific citations for patent market value but an insignificant effect on patent citations, and (6) in-text and front-page references yield substantially different analytical results. These results are robust when additionally controlling for the number of patent references and inventors (SM §4).

Within- vs. between-company effects. When we additionally incorporate fixed effects of the first applicant of the patent, results are noticeably different (SM §5). More specifically, based on in-text references, the positive effect of the number of references on market value (and partially patent citations as well) becomes insignificant, insignificant effects of interdisciplinarity on both patent citations and market value turn positive, insignificant effect of scientific citations on patent citations turns negative, and positive effect of scientific citations on market value turns insignificant. These suggest important differences between within-company and between-company effects. While companies using more science produce more valuable patents than companies using less science, within the same company, a particular technology using more science is not necessarily more valuable than another technology using less science, suggesting a spillover effect across different R&D projects within the same company. While companies using more interdisciplinary science do not produce more valuable patents than companies using less interdisciplinary science, within the same company, patents that use more interdisciplinary science is more valuable than patents using less interdisciplinary science. This might suggest a barrier for translating interdisciplinary science into valuable innovation at the company level, which however does not differ across different projects within the company. While companies using more highly cited science produce patents that are more valuable (especially in terms of the market value) than companies using less highly cited science, within the same company, patents using more highly cited science do not seem more valuable than patent using less highly cited science, again suggesting a spillover effect across different projects within the same company.

Level of IPC controls. Our regression analysis controls for IPC subclasses at the 4-digit level, and results are robust when controlling for IPC classes at the 3-digit level (SM §6).

Self-citations. We test the robustness of our results by excluding self-citations (SM §7). We find that 5.2% in-text references and 9.8% front-page references are self-citations. The shares are small and comparable to Marx and Fuegi (2022), 6 % and 10 % for in-text and front-page references, respectively. Our results regarding the relationship between science characteristics and patent value are consistent when excluding self-citations, except that the insignificant effect of scientific citations on patent citations becomes negatives. This suggests that citing others' highly cited science might lead to lower technological impact, while citing own highly cited science offsets this negative effect. This might be explained by the first-mover advantage for building on own highly cited science. Self-referenced papers in the patent text are more likely to be listed on the front page of the same patent, and our analytical results regarding which types of science are more likely to be listed on the front page are consistent when controlling for self-citations.

4. Discussion

We find that in-text references are more basic and have more scientific citations than front-page references. Marx and Fuegi (2022) also found that in-text references are more cited in science than front-page references but did not examine basicness. Using front-page references, prior studies have documented positive associations between the likelihood of being cited by patents and paper basicness (Ke, 2020), as well as science citations (Ahmadpoor & Jones, 2017; Hicks et al., 2000; Popp, 2017; Veugelers & Wang, 2019). These positive associations might be even underestimated considering that in-text references are more basic and highly cited in science than front-page references.

We also find that the difference between in-text and front-page references in interdisciplinarity and novelty is small when comparing at the reference level and insignificant when comparing at the patent level. Marx and Fuegi (2022) found that in-text references are more interdisciplinary than front-page references but did not examine novelty. Different results on interdisciplinarity might reflect different scientific domain are being analyzed: Marx and Fuegi (2022) examined all fields while we focus on the biotech domain. Prior studies documented that novel and interdisciplinary scientific papers are more likely to be cited by patents using front-page references (Ke, 2020; Veugelers & Wang, 2019), these findings might be robust if in-text references are being used instead.

We find a consistent positive association between patent forward citations and the number of science references, for both in-text and front-page ones. This is consistent with Bryan et al. (2020). We observe a similar pattern for patent market value, this is slightly different from Bryan et al. (2020), who found that having front-page references does not have an significant effect while the number of front-page references has a positive effect, and having in-text references has a positive effect while the number of in-text reference does not have a significant effect. This might be because Bryan et al. (2020) studies all technological fields while we focus on the biotech domain, which is known for being science-intensive. Another possible explanation is that Bryan et al. (2020) started from a set of journals and potentially missed in-text references to other journals.

More importantly, we find that using in-text and front-page references yield substantially different results regarding the relationship between patent value and characteristics of cited science in terms of basicness, interdisciplinarity, novelty, and scientific citations. These have not been examined in prior literature.

This study has several limitations. First, our dataset covers a sample of patents and all their references. However, we do not know whether the papers in our dataset are cited by patents outside our sample. This data limitation does not allow us to use scientific papers as the unit of analysis to investigate how different types of science are cited by all patents and how their technological impact unfolds. Second, we adopt a nonparametric approach (i.e., categorizing patents into deciles based on certain science measures) to uncover the complex associations between the characteristics of referenced science and patent value. This approach is simple but powerful for this exploratory research aiming at uncovering interesting patterns. Future research should develop more formal models to systematically explain these patterns. In particular, the complex association between novelty and patent value needs to be further investigated. Future research also needs to develop more sophisticated designs to allow causal inference. Third, results for patent citations are not identical to the result for patent market value. We need more research into the difference between these two measures, as well as their different response to the characteristics of referenced science. Fourth, the market value information is only available for patents from publicly listed companies, and therefore the findings about market value might not be generatable to other patents. Fifth, our analysis of the process through which in-text and front-page references are generated is very preliminary. To fully understand why front-page and in-text references lead to different analytical results, we need future research to further investigate the reference generation process, by collecting survey, interview, and archival data, and to design more rigorous and direct empirical tests. Sixth, our sample only covers USPTO biotech patents. Our findings might not be generalizable to other technological fields. How science and technology interact with each other may be different across fields. Our basicness measure is specific to the biomedical sciences but may not apply to other fields, such that the observed effects regarding basicness can only be interpreted in the context of biotech research and development. In addition, our findings may not apply to other patent authorities. Considering that the practice at USPTO is very different from that at EPO, as well as other patent offices, it is important to understand how the reference generating process might differ across patent authorities and how these differences might influence the validity of using patent references as a proxy of knowledge flow.

5. Conclusion

This paper assessed the differences between patent in-text and front-page references to science: (1) how they are different in terms of basicness, interdisciplinarity, novelty, and scientific citations, (2) whether they lead to the same results when analyzing the association between patent value and characteristics of referenced science, and (3) how different types of in-text references are more likely to be listed on the front page of the same patent. Using a dataset consisting of 33,337 USPTO biotech patents and their 860,879 in-text and 637,570 front-page references to Web of Science journal articles, we find that in-text references are more basic and highly cited in science than front-page references. The differences in terms of interdisciplinarity and novelty is small when comparing at the reference level but insignificant when comparing at the patent level. In terms of the relationship between patent value and science profile using in-text references, we found that patents that cite more scientific papers are more valuable than patents that cite fewer scientific papers, where patent value is measured by the number of citations a patent receives from future patents and the market value based on the stock market response to the issuing of the patent. Within the set of patents that cite science, the average basicness of referenced science has an inverted U-shaped relation with patent value. Interdisciplinarity does not have any significant association with patent citations or market value. The average scientific citations have an insignificant association with patent citations but a positive relation with patent market value. Relationship with novelty is complex. Patent citations have a sudden drop when moving from not citing any novel papers to citing novel papers, and then slowly rise to a higher level. On the other hand, patent market value has a sudden jump when moving from not citing any novel papers to citing novel papers, and then slowly declines. However, using front-page references cannot replicate all these results. Results are substantially different, especially for patent citations. We further examined what types of in-text referenced papers are more likely to be listed on the front page of the same patent. We found that papers are more likely to be listed on the front page when they are moderately basic, less interdisciplinary, less novel, and more highly cited in science.

This study contributes to the recurrent debate regarding the relationship between science and technology, more specifically, whether scientific research is useful for, or rather incompatible with, technological development. Our results suggest that both sides have merits and that unpacking the heterogeneity in scientific outputs provides a promising direction to reconcile these competing theories. Our results show that patent value responses to basicness, interdisciplinarity, novelty, and scientific citations in different ways. Our results also contribute to the studies of patent references to the scientific literature. The inconsistencies between the results based on in-text and front-page references warns that our analytical results might be sensitive to data source, and we need to be cautious about which type of references is more suitable for different types of questions that is being asked. For example, for tracing general knowledge flow, in-text references might be better suited as they capture various types of connections and knowledge spill-over. For mapping the intellectual structure underlying science and technology, front-page references might be better suited as they indicate stronger similarity between the citing patent the cited science, at least in the sense that front-page references are the ones that really challenge the novelty of the focal invention.

CRediT authorship contribution statement

Jian Wang: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Suzan Verberne:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Acknowledgements

This work was supported by a grant from the European Patent Office (EPO) - Academic Research Programme.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2024.101564.

References

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

Arts, S., Melluso, N., & Veugelers, R. (2023). Beyond citations: Measuring novel scientific ideas and their impact in publication text. *arXiv e-prints*, arXiv: 2309.16437.

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, 62*(10), 2765–2783. https://doi.org/10.1287/mnsc.2015.2285

Bryan, K. A., Ozcan, Y., & Sampat, B. (2020). In-text patent citations: A user's guide. *Research Policy, 49*(4), Article 103946. https://doi.org/10.1016/j.respol.2020.103946

Callaert, J., Pellens, M., & Van Looy, B. (2014). Sources of inspiration? Making sense of scientific references in patents. *Scientometrics, 98*(3), 1617–1629. https://doi.org/10.1007/s11192-013-1073-x

Callaert, J., Vervenne, J., Van Looy, B., Magermans, T., Song, X., & Jeuris, W. (2014). *Patterns of Science-Technology Linkage*. Retrieved from.

Fontana, M., Iori, M., Montobbio, F., & Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy, 49*(7), Article 104063.

Gibbons, M. (1994). *The new production of knowledge*. Sage Publ.

Gittelman, M., & Kogut, B. (2003). Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science, 49*(4), 366–382. https://doi.org/10.1287/mnsc.49.4.366.14420

Hamilton, K. (2003). *Subfield and level classification of journals (CHI No. 2012-R)*. Haddon Heights, NJ: CHI Research.

Hicks, D., Breitzman, A., Hamilton, K., Sr, & Narin, F (2000). Research excellence and patented innovation. *Science and Public Policy, 27*(5), 310–320. https://doi.org/10.3152/147154300781781805

Iori, M., Pelletier, P., Souza, D., Fontana, M., & Geuna, A. (2024). Mapping new knowledge: using novelty measures critically. In *Paper presented at the 17th Workshop On The Organisation*. Turin, Italy: Economics And Policy Of Scientific Research.

Ke, Q. (2020). Technological impact of biomedical research: The role of basicness and novelty. *Research Policy, 49*(7), Article 104071. https://doi.org/10.1016/j.respol.2020.104071

Ke, Q. (2023). Interdisciplinary research and technological impact: evidence from biomedicine. *Scientometrics*, 1–43.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth*. *The Quarterly Journal of Economics, 132*(2), 665–712. https://doi.org/10.1093/qje/qjw040

Marx, M., & Fuegi, A. (2020). *Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations*. Retrieved from.

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

Merton, R. K. (1973). *The sociology of science*. Chicago, IL: University of Chicago Press.

Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics, 49*(1), 93. https://doi.org/10.1023/a:1005613325648

Nagaoka, S., & Yamauchi, I. (2015). The use of science for inventions and its identification: patent level evidence matched with survey. *Research Institute of Economy, Trade and Industry (RIETI)*.

Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics, 7*(3), 369–381. https://doi.org/10.1007/BF02017155

Noma, E. (1986). *Subject classification and influence weights for 3,000 journals*. Cherry Hill, NJ: CHI Research/Computer Horizons.

Nunn, H., & Oppenheim, C. (1980). *A patent journal citation network on prostaglandins*. Elsevier.

Poege, F., Harhoff, D., Gaessler, F., & Baruffaldi, S. (2019). Science quality and the value of inventions. *Science Advances, 5*(12), eaay7323. https://doi.org/10.1126/sciadv.aay7323

Popp, D. (2017). From science to technology: The value of knowledge from different energy research institutions. *Research Policy, 46*(9), 1580–1594. https://doi.org/10.1016/j.respol.2017.07.011

Ramshaw, L. A., & Marcus, M. P. (1999). Text chunking using transformation-based learning. *Natural language processing using very large corpora* (pp. 157–176). Springer.

Sampat, B. (2010). When do applicants search for prior art? *The Journal of Law and Economics, 53*(2), 399–416. https://doi.org/10.1086/651959

Sang, E. T. K., & De Meulder, F. (2003). Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. In *Paper presented at the Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*.

Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature, 34*(3), 1199–1235. Retrieved from http://prx.library.gatech.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9610113170&site=ehost-live.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface, 4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Tamada, S., Naito, Y., Kodama, F., Gemba, K., & Suzuki, J. (2006). Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics, 68*, 289–302.

Tijssen, R. J. W., Buter, R. K., & van Leeuwen, T. N. (2000). Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics, 47*(2), 389–412. Retrieved from <Go to ISI>://WOS:000089449100014.

van Raan, A. F. J. (2017). Sleeping beauties cited in patents: Is there also a dormitory of inventions? *Scientometrics, 110*(3), 1123–1156. https://doi.org/10.1007/s11192-016-2215-8 van Raan, A. F. J., & Winnink, J. J. (2018). Do younger Sleeping Beauties prefer a technological prince? *Scientometrics, 114*(2), 701–717. https://doi.org/10.1007/s11192-017-2603-8

Verberne, S., Chios, I., & Wang, J. (2019). *Extracting and Matching Patent In-text References to Scientific Publications*. BIRNDL@ SIGIR. Paper presented at the.

Verluise, C., Cristelli, G., Higham, K., & de Rassenfosse, G. (2020). The missing 15 percent of patent citations. Available at SSRN 3754772.

Veugelers, R., & Wang, J. (2019). Scientific novelty and technological impact. *Research Policy, 48*(6), 1362–1372. https://doi.org/10.1016/j.respol.2019.01.019

Voskuil, K., & Verberne, S. (2021). Improving reference mining in patents with BERT. *arXiv preprint arXiv:2101.01039*.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics, 94*(3), 851–872.

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PloS one, 10*(5), Article e0127298.

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

Weber, G. M. (2013). Identifying translational science within the triangle of biomedicine. *Journal of Translational Medicine, 11*(1), 126. https://doi.org/10.1186/1479-5876-11-126

Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences* (2nd ed.). New York, NY: Oxford University Press. Oxford, UK.