Received: 15 November 2022 | Revised: 30 May 2025 | Accepted: 5 October 2025

DOI: 10.1002/smj.70027

RESEARCH ARTICLE

SMS | Strategic Management Journal

WILEY

Beyond the front page: In-text citations to patents as traces of inventor knowledge

Cyril Verluise[1] | Gabriele Cristelli[2] | Kyle Higham[3] | Gaétan de Rassenfosse[4]

[1]QuantumBlack, Paris, France

[2]Department of Geography and Environment, The London School of Economics and Political Science, London, UK

[3]Motu Economic and Public Policy Research, Wellington, New Zealand

[4]College of Management of Technology, École polytechnique fédérale de Lausanne, Lausanne, Vaud, Switzerland

Correspondence
Gaétan de Rassenfosse, College of Management of Technology, Ecole polytechnique fédérale de Lausanne, Station 5, 1015 Lausanne, Switzerland.
Email: gaetan.derassenfosse@epfl.ch

Abstract

Research Summary: This study introduces in-text patent-to-patent citations—references embedded in the body of patent documents—as a novel data source to trace knowledge flows. Unlike front-page citations, which often reflect legal requirements, in-text citations are more likely to originate from inventors and signal meaningful technological linkages. We show that they exhibit stronger geographic and semantic proximity, greater self-referentiality, and closer alignment with inventor knowledge. Though less frequent than front-page citations, they yield robust results in models of knowledge diffusion. We release a validated dataset and reproducible code to support future research. Our findings offer new opportunities for scholars interested in the microfoundations of innovation, the geography of knowledge flows, and the role of inventors in shaping firms' knowledge trajectories.

Managerial Summary: This study introduces in-text patent citations—references embedded in the technical description of patent documents—as a new way to trace how knowledge flows among inventors and firms. Unlike traditional front-page citations, which often reflect legal formalities or examiner input, in-text citations originate more directly from inventors and thus capture genuine technological linkages. Using large-scale U.S. patent data, we show that in-text citations connect geographically and thematically closer inventions and are more likely to involve the same firms or inventors. The dataset, released as open data, enables more accurate analyses of innovation dynamics and firm knowledge strategies. Managers can use these insights to better understand how ideas diffuse and where valuable inventive knowledge resides.

KEYWORDS

citation, knowledge flow, open data, patent, spillover

1 | INTRODUCTION

Innovation depends on how effectively firms generate, combine, and deploy knowledge, but the traces that these processes leave behind are faint. Despite its central role, knowledge remains notoriously challenging to observe. In this desert of data, patent data "loom up as a mirage of wonderful plentitude and objectivity" (Griliches, 1990, p. 1661). One prominent use of patent data was proposed by Jaffe et al. (1993), who established patent citations as a proxy for knowledge flows. Since then, a large body of strategic management scholarship has used patent citations to trace how ideas flow across organizations and inventors (e.g., Almeida & Kogut, 1999; Mowery et al., 1996; Rosenkopf & Nerkar, 2001; Singh, 2005; Singh & Marx, 2013).

Although patent citations are often viewed as an appealing proxy for the elusive "paper trail" of knowledge (Krugman, 1991), they suffer from well-documented shortcomings. Chief among them is that inventors exert only limited influence over which references appear on a patent's front page—the traditional source of patent citation data. Front-page citations emerge from a multi-stage and highly mediated process in which the examiner has the final say. The USPTO's post-2001 distinction between examiner- and applicant-supplied citations is a welcome advance (Thompson, 2006), yet it only partially resolves the problem. Some examiner citations were initially suggested by the applicant, and not all applicant citations have been provided by the actual inventors, leaving the true origin of the citation ambiguous.

A growing body of evidence confirms that front-page citations to patents only imperfectly capture knowledge flows. Alcácer et al. (2009) and Sampat (2010) show that "applicants" (including inventors) supply only a minority of patent citations and that citation behavior varies widely across firms. Lampe (2012) adds a strategic dimension, documenting deliberate withholding of material prior art—a pattern contested by Kuhn et al. (2023). Fadeev (2024) suggests that front-page patent citations are less about individual inventors' knowledge dissemination and more about the strategic relationships between firms. Adding to the noise, patent attorneys frequently insert citations on behalf of their clients (Jaffe et al., 2000; Wagner et al., 2014). Moreover, unlike citations in scientific papers, patent references serve a legal rather than a cognitive purpose (Meyer, 2000). Finally, front-page citations are only available for granted patents, leaving out close to half the patent applications that are withdrawn or abandoned (Carley et al., 2015). Together, these factors render front-page patent citations a notoriously noisy proxy for knowledge diffusion (Duguet & MacGarvie, 2005) and, some argue, a metric that may miss such flows altogether (Arora et al., 2018).

The limitations of front-page patent citations invite exploration of alternative indicators of knowledge flows. This paper examines in-text patent-to-patent citations—references that appear within the body of patent documents—as one such alternative. Although datasets containing these "internal" patent citations have been available for some time (Berkes, 2018, pre-1947; Marx & Fuegi, 2022; Verluise & de Rassenfosse, 2020), their distinctiveness and empirical value have not yet been systematically assessed.

Our analysis demonstrates that in-text patent-to-patent citations better reflect inventors' knowledge set than front-page citations. Descriptive evidence shows that in-text citations are (a) more geographically concentrated than front-page references—regardless of whether those references were added by examiners or applicants; (b) textually closer to the patents they cite, indicating tighter thematic proximity; and (c) substantially more self-referential, pointing back to the focal firm's earlier inventions. These patterns suggest that in-text citations are more likely to originate with inventors themselves rather than with patent attorneys, and thus offer a sharper lens on knowledge diffusion. Evidence from a survey of patent attorneys backs this claim: a randomly selected in-text patent-to-patent citation is 19%–44% more likely to originate with the inventor than is a randomly selected front-page "applicant" citation—the USPTO's catch-all label for any reference supplied under the applicant's name, whether by inventors, in-house counsel, or outside attorneys.

In addition to these analyses, we assess the data's practical utility by replicating Balsmeier et al. (2023). This study is one of the most recent and stringent studies of knowledge diffusion, leveraging variation from inventor death. Although in-text patent-to-patent citations are less numerous than front-page references, their relative scarcity need not be a limitation. The replication reveals that in-text citations yield statistically robust estimates, even when applied in such a restricted empirical setting.

We further systematically validate our in-text patent-to-patent citation dataset. Using an open-source machine-learning parser to locate references within the specification, we measure its precision and recall, document residual error patterns, and release the entire reproducible workflow. Our publicly available code cleans the parser's raw output, disambiguates citation strings, and maps them to standardized patent numbers—giving scholars a transparent, auditable foundation for future research. The dataset covers 49.5 million in-text patent citations matched to close to 8 million unique patent documents (with best coverage starting from 1976).

The rest of the paper is organized as follows. We start by providing background information on in-text patent-to-patent citations, highlighting their main differences with front-page citations. This section explains why inventors are more likely to exert more influence over the selection of in-text citations than over front-page citations. We then provide a statistical analysis of front-page vs. in-text citations. Next, we replicate a recent study using citation data to track knowledge flows. We conclude by offering avenues for future research. The details of the data creation pipeline and validation metrics are provided in the Appendix S1, together with the survey results and additional Supporting Information.

2 | THE EPISTEMOLOGY OF IN-TEXT CITATIONS

A US patent document is legally composed of three segments. The front page provides bibliographic data, the abstract, and the official list of cited prior art. It is followed, when required, by drawings that visually support the disclosure. The remainder constitutes the specification, which houses the full narrative of the invention—background, summary, detailed description—and the claims. Because patent offices digitize and release front-page citations, they have become the default raw material for empirical research. Yet, the specification itself contains additional references embedded in the text, which we refer to as in-text citations. These citations may point to any genre of prior art, notably earlier patents and scientific publications. Our study concentrates on the subset of references to the patent literature, noting that in-text citations to scholarly literature have been examined elsewhere (e.g., Bryan et al., 2020; Marx & Fuegi, 2022).

In-text citations are inserted to satisfy the statutory patentability criteria in US patent law. Applicants use them to (i) differentiate the invention from prior art to argue novelty (35 US Code §102) and non-obviousness (§103), (ii) demonstrate enablement under §112 by guiding readers to supporting technical detail, and (iii) illustrate utility (§101) through concrete applications (see, e.g., Barton, 2003; Feit, 2011). Because these rhetorical functions overlap only partly with the examination-oriented purpose of front-page references, in-text citations plausibly carry information not reflected in front-page citations. Further, we contend that this incremental signal is shaped by the inventor's input during specification drafting, making in-text patent citations a promising indicator of underlying knowledge flows (Bryan et al., 2020).

2.1 | A legal perspective on in-text patent citations

The justifications for adding in-text citations listed above map directly onto the statutory requirements that an application must satisfy to be patentable. While novelty and non-obviousness are primarily assessed by the examiner through direct comparison with prior art, enablement and utility are mainly argued by the applicant within the specification. Table 1 illustrates the variety of in-text citations and their (often explicit) legal purposes.

2.1.1 | Novelty and non-obviousness

Applicants disclose the prior art they know by filing one or more Information Disclosure Statements (IDSs), and most of these references later appear on the patent's front page.[1] That front-page list, however, does not include all citations in the IDSs, nor is it limited to those in the IDSs. Examiners run their own searches and routinely add references that were never in an IDS, marking them "cited by examiner." Conversely, an applicant-supplied reference may be retained yet re-labeled examiner-cited once the examiner confirms its relevance. The resulting front-page catalogue, therefore, blends multiple sources of prior art and only imperfectly represents the applicant's original knowledge set—even for the so-called "applicant" citations.

Applicants may also advance novelty and non-obviousness within the specification itself by contrasting the invention with specific prior patents. In that case, the same reference typically

[1]See 37 CFR §1.56 and 37 CFR §1.97.

TABLE 1 Typology of in-text citations.

Citation reason	Example patent	Citation and context
Enablement	9,607,299 (*Transactional security over a network*)	"Techniques for data encryption are disclosed in, for example, US Pat. Nos. 7,257,225 and 7,251,326 (incorporated herein by reference) and the details of such processes are not provided herein to maintain focus on the disclosed embodiments."
	9,606,907 (*Memory module with distributed data buffers and method of operation*)	"Examples of circuits which can serve as the control circuit … are described in more detail by US Pat. Nos. 7,289,386 and 7,532,537, each of which is incorporated in its entirety by reference herein."
Novelty and non-obviousness	8,100,652 (*Ceiling fan complete cover*)	"US Pat. No. 5,281,093, issued to Sedlak et al., discloses a fan blade cover with a zipper. Sedlak, however, does not protect the fan's housing and motor, nor does it prevent blades from spinning."
	9,607,328 (*Electronic content distribution and exchange system*)	"One skilled in the art will readily appreciate that there is a great deal of prior art centered on methods for selecting programming for a viewer based on previous viewing history and explicit preferences, for example, US Pat. No. 5,758,257. The methods described in this application are unique and novel over these techniques as they suggest…"
Usefulness	9,607,730 (*Non-oleic triglyceride based, low viscosity, high flash point dielectric fluids*)	Applicant directly compares empirical results for the invention at hand with similar, previously granted patents.

appears both in the text and on the front page. Note that some patent lawyers simply add any in-text citations onto the IDS just in case any are considered material to patentability.

2.1.2 | Enablement

Section 112 is the legal requirement that a patent's specification describes the invention clearly and completely enough for any skilled practitioner in the field to make and use it without undue experimentation. Applicants often streamline that task by incorporating earlier patents by reference, citing them in the specification. If these citations are not relevant to novelty or non-obviousness, applicants are not required to disclose them via an IDS. Therefore, these "enablement" citations are not necessarily duplicated on the front page of the patent document. This is particularly true of citations accompanying specific examples that describe how the invention may be used in practice ("best modes"), which may be complementary (and not necessarily similar) to the invention described and may even be hypothetical (Freilich, 2019).

2.1.3 | Utility

The invention must be "new and useful" to be patentable. The first part of this clause is covered by the novelty and non-obviousness requirements described above. The second, usually referred to as the "utility" requirement, is particularly open to interpretation, but generally requires the patented invention to work (Machin, 1999). Utility is often assumed, and rejections based on lack of utility are rare for most technology types, providing little incentive to add citations (Chien & Wu, 2018). While there is no burden on the applicant to prove that the invention works (Cotropia, 2009), applicants sometimes cite prior patents or scientific work to demonstrate that the claimed function is physically achievable. These references, like enablement citations, are more likely to remain confined to the specification.

2.2 | In-text patent citations are more likely to originate from inventors

Because in-text citations arise for a broader set of reasons than those that populate the IDS, they capture portions of the inventive knowledge base that never surface on the front page, particularly those items deemed necessary to meet the enablement or usefulness requirement.[2] Although in-text citations fulfill legal objectives that differ from scholarly attribution, inventor influence is nevertheless likely to be stronger here than on the front page for two reasons.

First, the in-text citations that are duplicated on the front page, as prior art material to patentability, are likely the most relevant pieces of prior art against which the invention needs to be judged as novel and non-obvious. The fact that these citations are also in the patent description would imply that they either fulfilled multiple requirements or were so technologically close to the citing patent that applicants need to make explicit arguments for novelty in the description with reference to specific items in the prior art. In either case, the inventor was likely aware of this art—or worked closely with counsel to frame the necessary technical distinctions during drafting.

Second, citations that remain only in the text likely serve utility or enablement functions. The enablement requirement states that a hypothetical "person skilled in the art" should be able to make and use the invention, and applicants add in-text citations to assist these hypothetical persons. Consequently, this information was almost certainly necessary during the invention process, and the inventors were, therefore, aware of it. It is difficult to imagine attorneys alone supplying such technical scaffolding without substantial input from the inventors themselves.

Both of these arguments imply that inventors are more likely to exert more influence over the selection of in-text citations than over front-page citations. Accordingly, we propose that in-text patent-to-patent citations constitute a promising indicator of knowledge flows.

3 | SEARCHING FOR TRACES OF INVENTOR'S INPUT

This section compares in-text and front-page patent citations along geographic, semantic, and bibliographic dimensions and summarizes survey evidence on citation provenance. The results

[2]See Manual of Patent Examining Procedure, Sections 2164 and 2107.02.

presented in this section collectively suggest that in-text citations to patents provide a stronger signal of knowledge flows than front-page citations. We begin with broad descriptive statistics for each citation type.

3.1 | Descriptive statistics

Table 2 describes the full sample of 16,781,144 US patents and patent applications in our dataset published between the first patent grant in 1790 and August 2019.[3] Of these, 7,869,894 (about 46.90%) contain at least one patent citation in the body of the specification, whereas 76.79% make at least one front-page patent citation.

In total, the full sample contains 49,542,360 in-text patent-to-patent citations, roughly one-fifth of the 265,659,106 front-page citations made by the same patents. Although the distribution of the number of in-text citations per patent is highly skewed, the unconditional mean is 2.95 in-text citations per patent, rising to 6.29 among patents that make at least one citation. For completeness, note that the 49.5 million figure relates to in-text citations that we were able to match to a DOCDB publication number. As further explained in the Appendix S1, we identified over 60 million traces of in-text patent citations, and we were able to associate 49.5 million of these with a standardized patent number.

The (unconditional) number of in-text patent-to-patent citations is of the same order of magnitude as that of in-text patent-to-article citations (3.51 in Bryan et al., 2020, table 1). However, front-page patent-to-patent citations are significantly more numerous than front-page patent-to-article citations (4.60 in Bryan et al., 2020, table 1 vs. 15.83 in our Table 2).

TABLE 2 In-text and front page citations summary statistics.

	Front page	In-text
Number of patents	16,781,144	16,781,144
Number of patents with at least one citation	12,887,079	7,869,894
Share of patents with at least one citation	76.79%	46.90%
Number of citations (DOCDB)	265,659,106	49,542,360
Average number of citations per patent	15.83	2.95
Average number of citations per patent—conditional on citing at least one patent	20.61	6.29
Median pairwise similarity (dot product) between citing and cited patent [lower quartile, upper quartile][a]	0.71 [0.62, 0.78]	0.80 [0.68, 0.88]
Share of cited patents in the same INPADOC family	1.63%	10.51%
Share of cited patents with at least one common inventor	5.98%	17.43%
Share of cited patents with at least one common assignee	9.26%	22.46%

[a]After removing within-DOCDB family citations (and after 1947 only).

[3]The data are available for both granted patents and patent applications. They can be accessed at https://doi.org/10.5281/zenodo.3710993. Additional information is provided on the project website: https://cverluise.github.io/PatCit/.

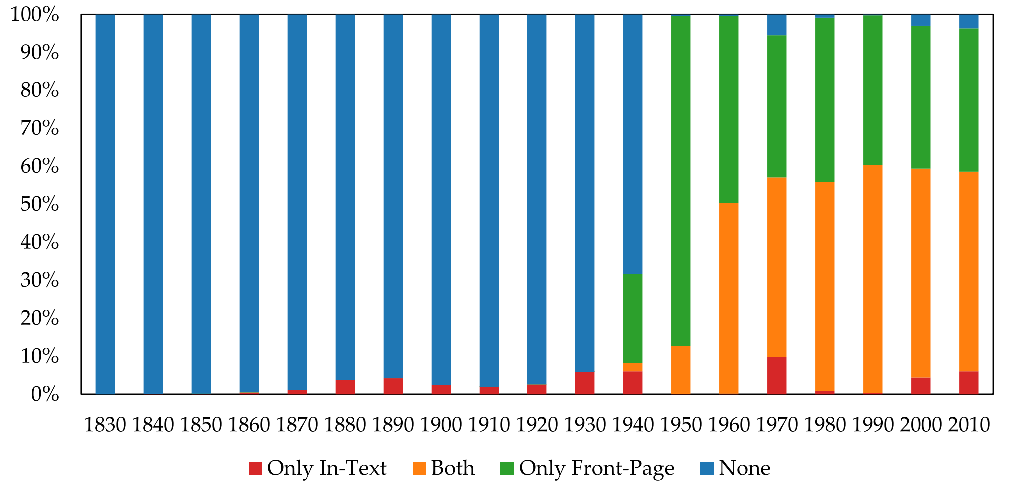

FIGURE 1 Citation types by decade. Proportion of patents by type of citation available.

We next examine the overlap between in-text and front-page citations. Figure 1 reports the proportion of citing patents by citation type. Before 1947, front-page citations did not exist; only in-text citations were available. However, fewer than 5% of patents from that era include in-text citations. In more recent decades, the share of patents with at least one in-text citation stabilizes around 60%, with most of these patents containing both in-text and front-page citations.

A notable shift occurs after November 2000, when the USPTO began publishing patent applications. This change allows researchers to observe the specification—including in-text citations—even for applications that are ultimately abandoned. As a result, the proportion of patent documents featuring only in-text citations increases during this later period.

Turning now to overlap in terms of citing–cited patent pairs, we assign each pair to one of three mutually exclusive sets: citation pairs found only in the text, only on the front page, or in both locations. We observe 11,799,723 pairs appearing in both places, accounting for just 5.79% of all front-page citations but 25.83% of all in-text citations. We observe that 33,883,406 in-text citations never appear on the front page, accounting for 14.25% of all 237,619,091 patent-to-patent citations recorded (not reported).

3.2 | In-text citations are more geographically concentrated than front-page citations

In this section, we compare the geographic properties of in-text and front-page citations. We begin by plotting the geographic distance between all pairs of citing and cited inventors' geocoded addresses using data from de Rassenfosse et al. (2019) for both in-text and front-page citations. We consider citing patents granted between 1980 and 2010 and exclude self-citations at the INPADOC family level.[4] Figure 2 shows the probability distribution functions of the distance, in kilometers, between citing and cited inventor dyads for the entire sample (panel A)

[4]INPADOC families group together all documents that share at least one priority filing, either directly or indirectly (e.g., via a third document). This "extended" definition contrasts with the narrower DOCDB definition (Martínez, 2011).

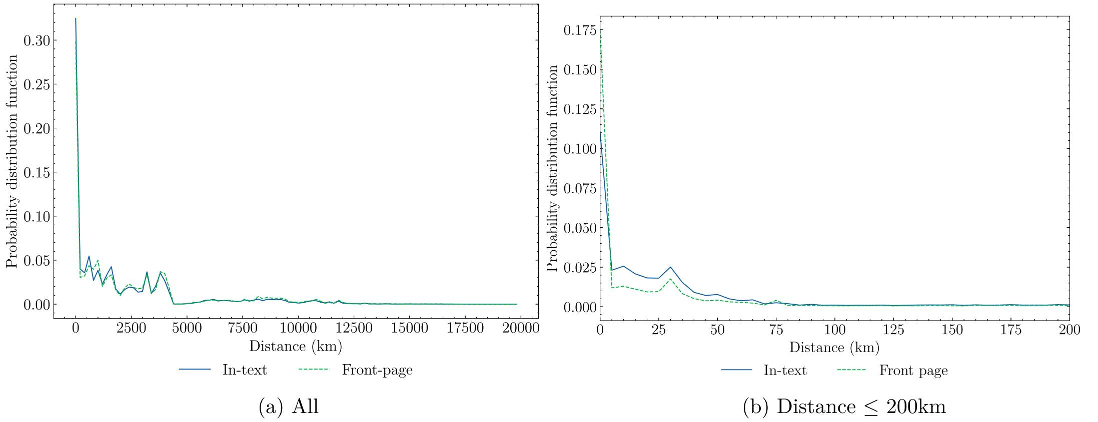

FIGURE 2 Citations distribution across cited inventors' location. Distance in kilometers is calculated from the latitude-longitude coordinates of the citing inventor's address to the latitude-longitude coordinates of cited inventor's address. The sample includes USPTO citing patents granted between 1980 and 2010. We exclude self citations at the INPADOC family level. In panel (a), we group observations by 200 km bins. In panel (b), we use 5 km bins.

and for citation pairs within 200 km (panel B). Both graphs portray in-text citations as slightly more localized than those on the front page. Panel B, in particular, shows a higher share of in-text citations within a 50-km distance.

We further compare in-text citations with two subgroups of front-page citations: those added by the applicant and those added by the examiner. The data for these comparisons are available for patents granted since 2001. Applicant front-page citations have been argued to be a less biased proxy of knowledge flows than examiner ones (Alcacer & Gittelman, 2006; Jaffe et al., 2000). As in Marx and Fuegi's (2022) study of patent-to-article citations, we adopt Thompson's (2006) approach, regressing a measure of geographic distance between citing and cited patents on a citation category indicator and citing patent fixed effects.

Table 3 reports our results. The sample includes patents granted between 2001 and 2010, and we continue to exclude self-citations at the INPADOC family level. In line with Thompson (2006), we find that applicant front-page citations are more localized than those added by the examiner. This result holds for both a coarse outcome, the probability of citing a patent originating from the same country (column 1, panel A), and a fine-grained outcome, the logarithmic transformation of the distance between citing and cited patents (column 1, panel B).

Comparing examiner and in-text citations (column 2), we find that in-text citations are also more localized than examiner front-page ones, and to a much greater extent than applicant front-page citations. Column (2) in panel A shows that in-text citations are, on average, 7 percentage points more likely to connect patents from the same country than examiner citations, whereas applicant citations are only about 2 percentage points more likely to do so. Column (2) in panel B indicates that in-text citations are, on average, approximately 67% closer than examiner front-page citations, compared to a difference of about 14% between applicant and examiner front-page citations.

Column (3) shows that in-text citations are also significantly more localized than applicant front-page citations, although the coefficients are, as expected, slightly smaller than those estimated in comparison with examiner citations. The economically and statistically significant

TABLE 3 Citations' geographic localization: In-text versus applicant- and examiner-front-page.

	Full sample			Restricted sample		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Same country						
Front-page App.	0.019			0.014		
	(0.000)			(0.001)		
In-text		0.073	0.042		0.050	0.023
		(0.001)	(0.000)		(0.001)	(0.001)
R^2	0.481	0.564	0.464	0.550	0.639	0.538
Observations	8,371,251	4,253,715	6,714,744	4,484,126	2,409,491	3,012,787
B. log Distance						
Front-page App.	−0.156			−0.124		
	(0.002)			(0.002)		
In-text		−1.102	−0.707		−0.602	−0.315
		(0.004)	(0.003)		(0.005)	(0.004)
R^2	0.295	0.412	0.356	0.458	0.405	0.711
Observations	8,371,251	4,253,715	6,714,744	4,484,126	2,409,491	3,012,787
C. log Distance (≤200 km)						
Front-page App.	0.017			−0.035		
	(0.004)			(0.007)		
In-text		−0.265	−0.225		−0.224	−0.134
		(0.006)	(0.004)		(0.014)	(0.007)
R^2	0.614	0.681	0.612	0.711	0.799	0.722
Observations	1,348,479	838,494	1,325,279	507,982	281,708	400,276
D. log Distance (within US)						
Front-page App.	−0.087			−0.079		
	(0.002)			(0.003)		
In-text		−1.040	−0.715		−0.547	−0.318
		(0.005)	(0.003)		(0.006)	(0.005)
R^2	0.286	0.413	0.352	0.335	0.448	0.398
Observations	6,218,864	3,051,426	5,147,672	3,257,424	1,696,826	2,259,050
Citing patent FE	✓	✓	✓	✓	✓	✓
Reference group	Front-page Exa.	Front-page Exa.	Front-page App.	Front-page Exa.	Front-page Exa.	Front-page App.

Note: Robust standard errors in parentheses. Estimations by OLS. Same country is a dummy variable equal to 1 for citing-cited pairs where the inventor countries coincide. Distance measures the kilometers separating the latitude-longitude coordinates of the citing and cited inventor. In the regressions, we employ its logarithmic transformation $log(1+distance)$. The sample includes USPTO citing patents granted between 2001 and 2010. We exclude self citations at the INPADOC family level. In columns 4, 5, and 6 we restrict the sample by (i) considering only citations between patents filed at up to 10 years of distance, (ii) excluding any patent with more than 100 front-page citations, (iii) excluding any self-citation at the patent applicant level. To identify unique patent applicants we use Du Plessis et al.'s (2009) identifiers.

greater localization of in-text citations persists when we restrict our sample to citation pairs within 200 km (panel C) or focus only on citation pairs within the United States (panel D).[5]

Our results are also robust to the use of restricted samples that exclude citations older than 10 years (column 4), patents with more than 100 front-page citations (column 5), and applicant self-citations (column 6).[6] The persistence of in-text citations' greater localization when we exclude applicant self-citations is particularly noteworthy, as it suggests that in-text citations may better capture technological knowledge flows not only within but also between organizations.

Our results consistently indicate that patent-to-patent in-text citations are decisively more localized than front-page ones. Just as the greater localization of applicant front-page citations relative to examiner ones may reflect the presence of fewer references unknown to the inventors, the strong geographic concentration of in-text citations—around the places where inventors work and live, and likely source a large share of their knowledge—suggests that in-text citations are a cleaner proxy for inventors' knowledge than front-page citations.

3.3 | In-text citations are textually more similar than front-page citations

To further assess the distinct nature of in-text patent-to-patent citations, we examine the textual similarity between citing and cited patents. Prior research has established that text-based similarity is a useful proxy for technological relatedness (e.g., Arts et al., 2021; Younge & Kuhn, 2016). In this context, a higher average similarity for in-text citations compared to front-page ones would suggest that they tend to connect more closely related inventions. Such a result would be consistent with the idea that in-text citations are more grounded in the substantive content of the invention and, therefore, more likely to reflect meaningful knowledge linkages.

We calculate the semantic similarity for a given patent pair as the dot product of Google Patents' document embedding vectors, which are made available to researchers through the Google Patents Public Datasets.[7] The embeddings are trained to predict CPC categories from each patent's full text using a WSABIE algorithm (Weston et al., 2010).

Table 2 shows that the median pairwise similarity between patents cited on the front page and the citing patent is 0.71. In contrast, the median similarity between patents cited in the specification and the citing patent is 0.80. Higher values indicate greater similarity, offering prima facie evidence that in-text citations connect conceptually closer patents than front-page citations.

Figure 3 plots the pairwise similarity distributions for in-text and front-page citations, alongside two reference distributions. The first reference distribution ("Within art unit") is based on the similarity between randomly chosen pairs of patents examined by the same USPTO art

[5]Interestingly, when restricting the sample to citations within 200 km, applicant front-page citations are *less* localized than examiner-added ones. However, the effect size is less than 2 percentage points.

[6]The results in columns (4–6) indicate a smaller difference between in-text and front-page citations than our baseline estimates, and a larger one between applicant and examiner front-page citations. Nevertheless, in-text citations remain substantially more localized than both groups of front-page citations.

[7]See https://tinyurl.com/googlepatentdata. A variety of similarity measures exist (see Ganguli et al., 2024, for a review); however, for the present work, we required a low-dimensional vector form that could quickly and intuitively estimate the semantic distance between a large set of patents. Google Patents' embeddings are perfect for this purpose.

unit.[8] The second reference distribution ('Random') is based on the similarity between in-text cited patents matched to a random citing patent. Specifically, we construct this set by randomly reassigning the cited patent in each in-text citation to another patent from the in-text citation pool, preserving the original set of citing and cited patents but randomly reconfiguring the citation links.

To ensure consistency, we restrict the analysis to citing and cited patents granted by the USPTO between 2000 and 2009 and exclude all within-INPADOC-family citations ($N = 325{,}247$). Self-family citations are much more frequent in the patent text than on the front page (as discussed in the next section); removing them improves the comparability of the similarity distributions. Pairs of patents used for the "Within art unit" and "Random" distributions are randomly sampled to match the sample size.

We draw two main observations from the graphical comparison. First, the modal peak of the in-text citation distribution is shifted toward higher similarity values compared to the front-page citation distribution. This shift indicates that patents cited in-text are, on average, more similar to the citing patent than those cited on the front page. Such a pattern is consistent with the idea that inventors reference closely related prior art in the specification—often to distinguish their invention from it or explain its relevance, as explained above. This pattern is also consistent with applicants (and attorneys) erring on the side of caution by over-disclosing

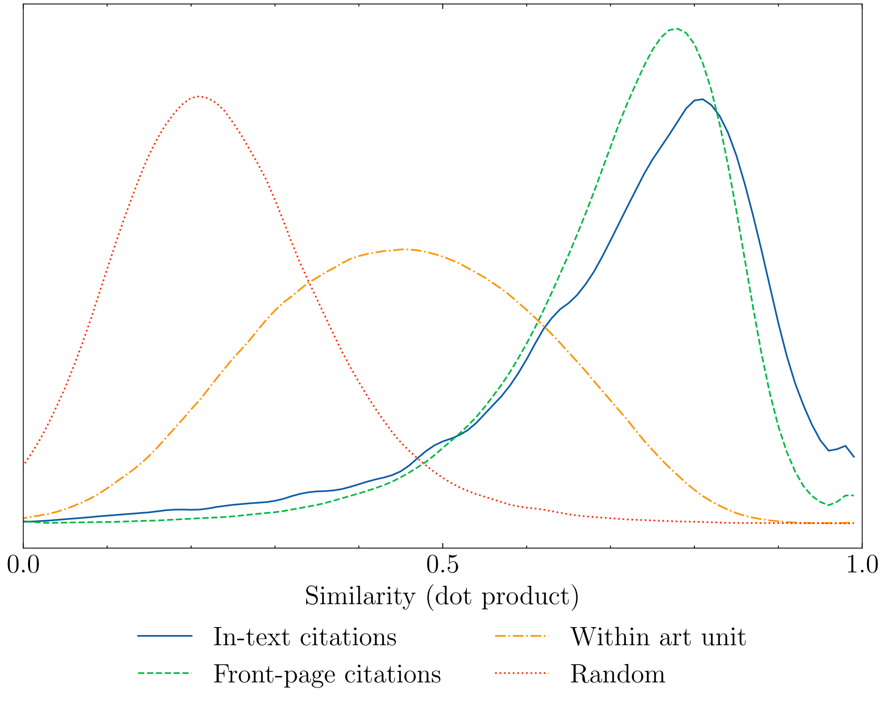

FIGURE 3 Citing-cited patent pairwise similarity distribution. Within-INPADOC-family citations omitted.

[8]An art unit is a group of patent examiners organized around specific technology areas. There are more than 600 art units at the USPTO.

known prior art on IDSs and may also reflect a deliberate attempt to overwhelm the patent office to hide actual relevant references (Bryan et al., 2020; Taylor, 2012).

Second, the in-text citation distribution displays a heavier left tail, with more citations at lower similarity levels, particularly around the level observed among patents examined by the same art unit. This is expected: in-text citations are not constrained by legal relevance for patentability and may include prior art from a broader, yet still related, technological space. Importantly, Figure 3 suggests that these citations constitute a small minority of in-text citations, at least to the extent that semantic similarity can measure this kind of relationship.

Taken together, these findings reinforce the view that in-text citations are more tightly coupled to the technical content of the citing invention than front-page citations. While not immune to noise, they appear to offer a more selective and content-driven signal of knowledge connection, supporting their potential as a valuable proxy for tracing knowledge flows.

3.4 | In-text citations are more self-referential than front-page citations

To further understand the origins and informational content of in-text citations, we examine their degree of self-referentiality. A higher prevalence of self-citations among in-text references, compared to front-page ones, may reinforce the view that they are more tightly linked to the knowledge base actively used by inventors, rather than serving purely legal or administrative functions.

We consider two forms of self-citations: family-level self-citations and applicant-level self-citations. To assess family-level self-citations, we map each citing and cited patent to its corresponding INPADOC patent family and compute the share of citations in which the cited patent belongs to the same family as the citing one.[9] We find that 10.51% of in-text citations are family self-citations, whereas the corresponding value for front-page citations is 1.63% (see Table 2). This source of information is not particularly informative—the patent is usually citing an earlier version of itself, typically in the "related applications" section of the description. From a practical standpoint, these citations are easy to identify and omit. Importantly, they represent only a small share of in-text citations overall.

Turning to applicant-level self-citations, we compute the share of citations in which the citing and cited patents share at least one inventor or one assignee. We rely on harmonized names provided in the Google Patents dataset, labeling a citation as "same-applicant" when at least one inventor or assignee name is shared between the citing and cited patents. This approach casts a wide net: inventors may move between firms, and the name harmonization process is more likely to merge distinct entities (due to common names or errors) than to split identical ones. As a result, our estimates likely overstate the true number of applicant-level self-citations. However, this issue affects front-page and in-text citations equally, limiting concerns about its impact on the comparative analysis.

The differences we observe are substantial. Among in-text citations, 17.43% share at least one inventor and 22.46% share at least one assignee with the citing patent (see Table 2). For front-page citations, the corresponding figures are 5.98% and 9.26%. These findings highlight the importance of self-reliance in knowledge creation and suggest that in-text citations more

[9]Performing the analysis on the DOCDB family leads to substantially similar conclusions.

prominently capture this phenomenon. More broadly, they provide further evidence that in-text citations more directly reflect the knowledge set of inventors than do front-page citations.

Self-citations are a foundational tool for measuring knowledge flows in strategic management research, being commonly used to capture knowledge reuse (e.g., Berry, 2014; Melero et al., 2020) or to operationalize constructs such as generative appropriability (Ahuja et al., 2013; Argyres et al., 2025). Scholars have employed variations of self-citation-based metrics to examine knowledge transfer across a firm's divisions (Miller et al., 2007), knowledge "inheritance" from prior employers (Chatterji, 2009), and the "specificity" of a firm's knowledge base (Wang et al., 2009, 2016), among other applications. Our findings call for greater attention to in-text citations as a lens for analyzing intra-organizational knowledge dynamics.

3.5 | Patent attorneys believe that in-text citations are more likely to originate from inventors

To complement our empirical analyses, we conducted a survey of US patent attorneys to gain insight into the likely origin of citations found in patent applications.[10] Attorneys are uniquely positioned to answer this question: they directly observe the references they add and those that the applicant supplied. In the latter case, they may also directly exchange information with inventors, allowing them to observe the entire chain of events.

The survey focused on two types of citations: those listed on the IDS filings (from which "applicant" citations are drawn) and those embedded in the body of the specification (in-text). Respondents were asked to estimate how often each type of citation is supplied by the applicant (vs. themselves) and, among those supplied by applicants, how frequently the inventor is the source.

The results suggest that in-text citations are significantly more likely than front-page citations to originate with inventors. According to respondents, approximately 29%–38% of in-text patent citations are inventor-supplied, compared to 20%–29% for citations listed in the IDSs. This implies that a randomly selected in-text citation to a patent is between 19% and 44% more likely to have been provided by the inventor than a comparable front-page applicant citation. (See Table D14 in Appendix S1 for details.)

The survey also asked about scientific references. Respondents believe that inventors supply around 54%–57% of in-text citations to scientific papers, compared to 50%–57% for IDS-listed references. We further estimate that a randomly selected in-text citation to a scientific article is between 10 and 17% more likely to have been provided by the inventor than a comparable front-page applicant citation. The strength of the "inventor signal" is weaker for in-text articles than for in-text patents relative to front-page references. However, according to survey respondents, in-text and front-page scientific references are more likely to be suggested by the inventor than patent references.

These findings provide additional evidence that in-text patent-to-patent citations better reflect the knowledge set that inventors actively draw upon during the invention process, compared to front-page patent-to-patent citations. While attorneys often add references to comply with patentability standards, citations embedded in the technical narrative are more likely to be anchored in the inventor's own understanding of the relevant prior art.

[10]We refer readers to Section D of the Appendix S1 for a complete account of the survey methodology, sample characteristics, and detailed estimation procedures.

4 | DO IN-TEXT CITATIONS PROVIDE A STRONG ENOUGH SIGNAL OF KNOWLEDGE FLOWS?

The previous section established that in-text citations are more likely to originate from inventors than front-page citations. Accordingly, they provide a clearer signal of knowledge flows than the citation data scholars have traditionally relied upon. A potential limitation, however, is their relative sparseness compared to front-page citations, as shown in Table 2. The signal may be cleaner, but is it strong enough to be practically useful? For narrowly focused studies that aim to rigorously assess the use of citations as indicators of knowledge diffusion, the answer might well be no. To explore this possibility, we replicate one of the most recent and methodologically demanding studies in this area: Balsmeier et al. (2023).

4.1 | Empirical approach

A large body of research on the localization of knowledge spillovers builds on the matching approach developed by Jaffe et al. (1993). Starting from a given sample of patents, this approach involves matching patents that cite the focal sample with patents that are similar in application date and technology field, and estimating differences in the two citing populations' geographic locations. While very influential, this approach has been subject to a number of criticisms, as extensively discussed by Thompson and Fox-Kean (2005).

Balsmeier et al. (2023) propose an alternative identification strategy based on co-invented patents in which one of the inventors died after the application date but before the patent was granted. They identify 1621 such patents filed between 1976 and 2005, drawn from a universe of approximately 3 million granted patents over that period. These patents involve 1621 deceased inventors and 3870 surviving co-inventors, all located in geographically dispersed areas. Although the resulting sample is highly selective, the setting enables a natural experiment: because the deceased and surviving inventors contributed to the same invention, their relative influence on subsequent citations can be cleanly compared.

To measure local knowledge spillovers, the authors track the geographic distance between the hometowns of inventors on citing patents and those of the cited inventors—both deceased and surviving. They then compare the volume of citations originating from within concentric distance bands around each inventor's hometown. Because the cited invention is held constant across all observations, any systematic reduction in citations near the deceased inventor's location—relative to that of the surviving co-inventors—can be attributed to the loss of that individual's contribution to knowledge diffusion.

4.2 | Results

Using the original replication code, we recover 34,586 front-page citations (slightly fewer than the 34,749 reported in the original study), corresponding to 28,398 citing patents and 1621 cited patents. In contrast, we identify 6360 in-text citations, representing 5605 citing patents and 988 cited patents. Despite the smaller sample, there is no evidence of bias in the timing of citations: the distribution of citation ages is nearly identical across front-page and in-text citations (not reported).

However, the spatial distribution of citations differs substantially. In-text citations are notably more localized. For instance, while the original study reports that 15% of front-page citations occur within 10 miles of the cited inventor, our replication using in-text citations yields a corresponding figure of 22%. This eight-percentage-point difference persists across distances under 150 miles (not reported). Given our earlier findings on the geographic concentration of in-text citations, this result does not come as a surprise.

Table 4 compares the original results from Table 2 in Balsmeier et al. (2023) with our replications using both front-page and in-text citations. Our replication with front-page citations closely matches the original estimates, though not exactly.

For the full sample, front-page citations fall by 15%–25% within a 30-mile radius of the deceased inventor's location, with the effect diminishing to approximately 6.5% at 60 miles before losing statistical significance. In contrast, in-text citations fall by 9%–15% within 30 miles, tapering to about 5.5% at 60 miles, where they also lose significance.

In the long-distance sample, in which the deceased inventor lived at least 500 miles from all surviving co-inventors, front-page citations fall by 63%–76% within 30 miles and by 39% at 150 miles (remaining statistically significant). In-text citations in this group fall by 64%–73% within 30 miles, dropping to 22% at 150 miles, a result that is no longer statistically significant.

Overall, the findings are similar across front-page and in-text citations, with two notable differences. First, the drop in in-text citations within 20 miles in the main sample is smaller than the corresponding drop in front-page citations. This is consistent with the idea that applicant-submitted front-page citations often include a broader set of useful but non-essential references,

TABLE 4 Localization of knowledge flows.

	Cites from within X miles						
	10	**20**	**30**	**40**	**50**	**100**	**150**
Main sample							
Original values—front-page	−0.246	−0.299	−0.190	−0.101	−0.072	−0.016	−0.031
	(0.080)	(0.065)	(0.045)	(0.031)	(0.030)	(0.028)	(0.025)
Replicated values—front-page	−0.238	−0.292	−0.185	−0.097	−0.070	−0.013	−0.027
	(0.079)	(0.064)	(0.046)	(0.031)	(0.029)	(0.028)	(0.025)
Replicated values—in-text	−0.098	−0.139	−0.160	−0.074	−0.063	−0.007	−0.021
	(0.105)	(0.074)	(0.063)	(0.032)	(0.026)	(0.030)	(0.027)
Large distance sample							
Original values—front-page	−1.391	−1.225	−0.997	−0.954	−0.804	−0.604	−0.512
	(0.287)	(0.257)	(0.234)	(0.234)	(0.218)	(0.210)	(0.208)
Replicated values—front-page	−1.402	−1.231	−1.011	−0.969	−0.815	−0.607	−0.492
	(0.290)	(0.261)	(0.237)	(0.236)	(0.220)	(0.211)	(0.215)
Replicate values—in-text	−1.070	−1.293	−1.041	−1.042	−0.971	−0.410	−0.246
	(0.466)	(0.468)	(0.439)	(0.421)	(0.393)	(0.643)	(0.672)

Note: Estimates are based on a Poisson regression model. Standard errors in parentheses. The "large distance" sample refers to cases where all co-inventors are located at least 500 miles apart. We report Panels A and C from the original study. Consistent with the original findings, we find no significant effects of premature deaths when using in-text citation data (Panel B in the original study) and omit these estimates for readability.

which may be more marginal to the inventive contribution. When a co-inventor passes away, the more peripheral front-page citations—which may rely more on informal local pointers—drop more sharply.

Second, in-text citations in the large distance sample exhibit a smaller drop beyond the 100-mile mark compared to front-page citations. The original study finds a significant long-distance effect for front-page citations, which is somewhat surprising. The lack of a comparable effect for in-text citations aligns more closely with prior expectations about the localized nature of knowledge spillovers.

5 | CONCLUDING REMARKS

This study introduces and analyzes in-text patent-to-patent citations as a promising data source for examining knowledge flows. By distinguishing these citations from traditional front-page references, we document that in-text citations carry a stronger inventor-driven signal of technological influence. This finding suggests that in-text citations can serve as a valuable tool for scholars seeking to understand the microfoundations of innovation (e.g., Felin et al., 2015; Grigoriou & Rothaermel, 2014), intra-organizational knowledge dynamics (e.g., Miller et al., 2007; Wang et al., 2009), and the strategic behavior of firms in the patenting process (e.g., Ceccagnoli, 2009; Somaya, 2012). We highlight five broad avenues for future research.

First, in-text citations are available for abandoned (and published) applications—a feature that front-page citations lack. Accordingly, they open new possibilities for studying, for example, the antecedents of invention novelty, and lack thereof (e.g., Chen et al., 2021; Jung & Lee, 2016), and the decision-making processes behind patent abandonment (Somaya, 2012). Furthermore, these data provide a window into applicant citations before 2000. All studies from before this time can now be reassessed using this information, particularly if they focus on historical contexts such as the impact of the Bayh-Dole Act (e.g., Sampat, 2006) or the introduction of software patents (e.g., Hall & MacGarvie, 2010).

Second, researchers can leverage the fact that in-text patent-to-patent citations offer a cleaner lens on knowledge flows than traditional front-page references. Because earlier studies relied on noisier data, many promising research efforts may have succumbed to the well-known "file-drawer problem," where null or ambiguous results remain unpublished (Rosenthal, 1979), potentially distorting our understanding of knowledge diffusion. Our findings suggest that some of these shelved analyses—or hypotheses not supported by the data—may yield significant effects if re-estimated using in-text citations. Therefore, we encourage scholars to revisit abandoned paths and capitalize on this sharper signal to strengthen the empirical foundations of innovation research.

Third, while this study has focused on citations as proxies for knowledge flows, patent citations have long been used to measure a broader range of constructs, including most notably patent "importance" (Jaffe & de Rassenfosse, 2017). Definitions of "importance" vary across contexts (capturing economic value, technical merit, or legal robustness), and scholars readily acknowledge the limitations of existing patent metrics. Yet, the absence of viable alternatives has led to an over-reliance on front-page citations. In-text citations could be a helpful complement. A systematic exploration of in-text patent-to-patent citations as indicators of patent impact is a promising line of inquiry, enabling improvements in the precision of long-standing measures such as patent generality, breakthroughs, or patent centrality (Ahuja & Morris Lampert, 2001; Gilsing et al., 2008; Trajtenberg et al., 1997)—or producing entirely new measures.

Fourth, in a related vein, much remains to be understood about the relationship between in-text and front-page citations, especially where they overlap. For example, when an in-text citation also appears on the front page—particularly when added by an examiner—it may signal especially high relevance. Such overlap implies that both the applicant (perhaps even the inventor) and the examiner deemed the prior art pertinent, whether to satisfy disclosure obligations, frame novelty claims, or evaluate patentability. If these cited patents originate from firms operating in similar technological domains, they may offer a rich signal of competitive positioning or technological rivalry (Arts et al., 2025; McGahan & Silverman, 2006).

Finally, from a data perspective, future work could focus on extracting the context in which in-text citations appear. Not all references serve the same legal or rhetorical purpose, and recent work in bibliometrics has shown the value of analyzing citation contexts in scientific literature (e.g., Jurgens et al., 2018; Nicholson et al., 2021). We see strong potential for in-text data to enrich research on established phenomena where citation context matters. Because in-text citations embed more of the inventor's voice, they offer a window into the cognitive underpinnings of inventive activity, helping advance strategic management's broader ambition to understand the microfoundations of innovation.

ACKNOWLEDGMENTS

We thank the editor and two anonymous reviewers for their constructive feedback. We are also grateful to Philippe Aghion, Antonin Bergeaud, Bronwyn Hall, Adam Jaffe, and Francesco Lissoni for helpful comments, and to Patrice Lopez (WIPO) and Ian Wetherbee (Google) for their technical support. Lucas Violon provided outstanding research assistance. All remaining errors are ours. Open access publishing facilitated by Ecole polytechnique federale de Lausanne, as part of the Wiley - Ecole polytechnique federale de Lausanne agreement via the Consortium Of Swiss Academic Libraries.

DATA AVAILABILITY STATEMENT

All data and code are available and thoroughly described in the Online Appendix.

ORCID

Gaétan de Rassenfosse https://orcid.org/0000-0002-7862-0918

REFERENCES

Ahuja, G., Lampert, C. M., & Novelli, E. (2013). The second face of appropriability: Generative appropriability and its determinants. *Academy of Management Review*, *38*(2), 248–269.

Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6–7), 521–543.

Alcacer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics*, *88*(4), 774–779.

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in us patents: An overview and analysis. *Research Policy*, *38*(2), 415–427.

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, *45*(7), 905–917.

Argyres, N., Rios, L. A., & Silverman, B. S. (2025). On the heels of giants: Internal network structure and the race to build on prior innovation. *Strategic Management Journal*, *46*, 1728–1761.

Arora, A., Belenzon, S., & Lee, H. (2018). Reversed citations and the localization of knowledge spillovers. *Journal of Economic Geography*, *18*(3), 495–521.

Arts, S., Cassiman, B., & Hou, J. (2025). Technology differentiation, product market rivalry, and m&a transactions. *Strategic Management Journal*, *46*, 837–862.

Arts, S., Hou, J., & Gomez, J. C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, *50*(2), 104144.

Balsmeier, B., Fleming, L., & Lück, S. (2023). Isolating personal knowledge spillovers: Coinventor deaths and spatial citation differentials. *American Economic Review: Insights*, *5*(1), 21–33.

Barton, J. H. (2003). Non-obviousness. *IDEA: The Journal of Law and Technology*, *43*, 475–508.

Berkes, E. (2018). Comprehensive universe of US patents (CUSP): Data and facts. *Unpublished manuscript*, Ohio State University.

Berry, H. (2014). Global integration and innovation: multicountry knowledge generation within mnc s. *Strategic Management Journal*, *35*(6), 869–890.

Bryan, K. A., Ozcan, Y., & Sampat, B. (2020). In-text patent citations: A user's guide. *Research Policy*, *49*(4), 103946.

Carley, M., Hedge, D., & Marco, A. (2015). What is the probability of receiving a us patent. *Yale Journal of Law & Technology*, *17*, 203–223.

Ceccagnoli, M. (2009). Appropriability, preemption, and firm performance. *Strategic Management Journal*, *30*(1), 81–98.

Chatterji, A. K. (2009). Spawned with a silver spoon? entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal*, *30*(2), 185–206.

Chen, T., Kim, C., & Miceli, K. A. (2021). The emergence of new knowledge: The case of zero-reference patents. *Strategic Entrepreneurship Journal*, *15*(1), 49–72.

Chien, C. V., & Wu, J. Y. (2018). Decoding patentable subject matter. *Patently-O Patent Law Journal*, *1*(1), 10–19.

Cotropia, C. A. (2009). The folly of early filing in patent law. *Hastings Law Journal*, *61*, 65–129.

de Rassenfosse, G., Kozak, J., & Seliger, F. (2019). Geocoding of worldwide patent data. *Scientific Data*, *6*(1), 260.

Du Plessis, M., Looy, B. V., Song, X., & Magerman, T. (2009). Data production methods for harmonized patent indicators: Assignee sector allocation. EUROSTAT Working Paper and Studies.

Duguet, E., & MacGarvie, M. (2005). How well do patent citations measure flows of technology? Evidence from french innovation surveys. *Economics of Innovation and New Technology*, *14*(5), 375–393.

Fadeev, E. (2024). Creative construction: Knowledge sharing and cooperation between firms. Technical report, Working paper, Duke University.

Feit, I. N. (2011). Does a utility that is unproved at the time of filing violate sec. 112. *Journal of the Patent & Trademark Office Society*, *93*, 1–18.

Felin, T., Foss, N. J., & Ployhart, R. E. (2015). The microfoundations movement in strategy and organization theory. *Academy of Management Annals*, *9*(1), 575–632.

Freilich, J. (2019). Prophetic patents. *UC Davis Law Review*, *53*, 663–731.

Ganguli, I., Lin, J., Meursault, V., & Reynolds, N. F. (2024). Patent text and long-run innovation dynamics: The critical role of model selection. NBER Working Paper, (w32934).

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & Van Den Oord, A. (2008). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy*, *37*(10), 1717–1731.

Grigoriou, K., & Rothaermel, F. T. (2014). Structural microfoundations of innovation: The role of relational stars. *Journal of Management*, *40*(2), 586–615.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*(4), 1661–1707.

Hall, B. H., & MacGarvie, M. (2010). The private value of software patents. *Research Policy*, *39*(7), 994–1009.

Jaffe, A. B., & de Rassenfosse, G. (2017). Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*, *68*(6), 1360–1374.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, *90*(2), 215–218.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Jung, H. J., & Lee, J. J. (2016). The quest for originality: A new typology of knowledge search and breakthrough inventions. *Academy of Management Journal*, *59*(5), 1725–1753.

Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, *6*, 391–406.

Krugman, P. R. (1991). *Geography and trade*. MIT Press.
Kuhn, J., Younge, K., & Marco, A. (2023). Strategic citation: A reassessment. *Review of Economics and Statistics*, *105*(2), 458–466.
Lampe, R. (2012). Strategic citation. *Review of Economics and Statistics*, *94*(1), 320–333.
Machin, N. (1999). Prospective utility: A new interpretation of the utility requirement of section 101 of the patent act. *California Law Review*, *87*, 421–456.
Martínez, C. (2011). Patent families: When do different definitions really matter? *Scientometrics*, *86*(1), 39–63.
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.
McGahan, A. M., & Silverman, B. S. (2006). Profiting from technological innovation by others: The effect of competitor patenting on firm value. *Research Policy*, *35*(8), 1222–1242.
Melero, E., Palomeras, N., & Wehrheim, D. (2020). The effect of patent protection on inventor mobility. *Management Science*, *66*(12), 5485–5504.
Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics*, *49*(1), 93–123.
Miller, D. J., Fern, M. J., & Cardinal, L. B. (2007). The use of knowledge for technological innovation within diversified firms. *Academy of Management Journal*, *50*(2), 307–325.
Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1996). Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal*, *17*(S2), 77–91.
Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N. P., Grabitz, P., & Rife, S. C. (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*, *2*(3), 882–898.
Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, *22*(4), 287–306.
Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.
Sampat, B. N. (2006). Patenting and us academic research in the 20th century: The world before and after bayh-dole. *Research Policy*, *35*(6), 772–789.
Sampat, B. N. (2010). When do applicants search for prior art? *The Journal of Law and Economics*, *53*(2), 399–416.
Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, *51*(5), 756–770.
Singh, J., & Marx, M. (2013). Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity. *Management Science*, *59*(9), 2056–2078.
Somaya, D. (2012). Patent strategy and management: An integrative review and research agenda. *Journal of Management*, *38*(4), 1084–1114.
Taylor, R. B. (2012). Burying. *Michigan Telecommunications and Technology Law Review*, *19*, 99–130.
Thompson, P. (2006). Patent citations and the geography of knowledge spillovers: Evidence from inventor-and examiner-added citations. *Review of Economics and Statistics*, *88*(2), 383–388.
Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, *95*(1), 450–460.
Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, *5*(1), 19–50.
Verluise, C., & de Rassenfosse, G. (2020). Patcit: A comprehensive dataset of patent citations [data set]. Zenodo. https://doi.org/10.5281/zenodo.3710993
Wagner, S., Hoisl, K., & Thoma, G. (2014). Overcoming localization of knowledge? The role of professional service firms. *Strategic Management Journal*, *35*(11), 1671–1688.
Wang, H., He, J., & Mahoney, J. T. (2009). Firm-specific knowledge resources and competitive advantage: the roles of economic-and relationship-based employee governance mechanisms. *Strategic Management Journal*, *30*(12), 1265–1285.
Wang, H., Zhao, S., & He, J. (2016). Increase in takeover protection and firm knowledge accumulation strategy. *Strategic Management Journal*, *37*(12), 2393–2412.

Weston, J., Bengio, S., & Usunier, N. (2010). WSABIE: Scaling up to large vocabulary image annotation. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Three*, pages 2764–2770.

Younge, K. A., & Kuhn, J. M. (2016). Patent-to-patent similarity: A vector space model. Available at SSRN. https://ssrn.com/abstract=2709238

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Verluise, C., Cristelli, G., Higham, K., & de Rassenfosse, G. (2025). Beyond the front page: In-text citations to patents as traces of inventor knowledge. *Strategic Management Journal*, 1–21. https://doi.org/10.1002/smj.70027