Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 17-44
Revised September 2019
December 2017
https://doi.org/10.21799/frbp.wp.2017.44

The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences

Ina Ganguli
University of Massachusetts–Amherst

Jeffrey Lin
Federal Reserve Bank of Philadelphia Research Department

Nicholas Reynolds
Brown University

ISSN: 1962-5361

The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences*

Ina Ganguli† Jeffrey Lin‡ Nicholas Reynolds§

September 2019

Abstract

We show evidence of localized knowledge spillovers using a new database of U.S. patent interferences terminated between 1998 and 2014. Interferences resulted when two or more independent parties submitted identical claims of invention nearly simultaneously. Following the idea that inventors of identical inventions share common knowledge inputs, interferences provide a new method for measuring knowledge spillovers. Interfering inventors are 1.4 to 4 times more likely to live in the same local area than matched control pairs of inventors. They are also more geographically concentrated than citation-linked inventors. Our results emphasize geographic distance as a barrier to tacit knowledge flows.

Keywords: Localized knowledge spillovers, multiple invention, patents, interferences
JEL classification: O30, R12

*We thank Marcus Berliant, Jerry Carlino, Bob Hunt, and John Stevens, and conference and workshop participants at the Urban Economics Association, the Federal Reserve System Committee on Regional Analysis, the NBER Productivity workshop, EFPL Lausanne, the HBS Entrepreneurial Management seminar, the CID Growth Lab, George Washington University, and the Twelfth Annual Conference on Innovation Economics at Northwestern for comments and suggestions, and Aaron Rosenbaum for excellent research assistance.

†Department of Economics, University of Massachusetts Amherst, `iganguli@econs.umass.edu`.
‡Research Department, Federal Reserve Bank of Philadelphia, `jeff.lin@phil.frb.org`.
§Department of Economics and Population Studies and Training Center, Brown University, `nicholas_reynolds@brown.edu`.

Why is innovative activity geographically clustered (Carlino and Kerr, 2015; Lin, 2011)? One intriguing hypothesis is that inventors benefit from *localized knowledge spillovers*. Agglomeration may increase the frequency of interactions and the spread of knowledge, especially tacit or difficult-to-codify knowledge (Marshall, 1920; Glaeser, 1999; Feldman, 2000; Ganguli, 2015). The identification of localized knowledge spillovers faces at least two challenges. One, knowledge spillovers are hard to measure: "They leave no paper trail by which they may be measured or tracked" (Krugman, 1991). Two, different theories of agglomeration are "observationally equivalent," i.e. they yield similar predictions for measured productivity, wages, or other aggregates (Duranton and Puga, 2004; Audretsch and Feldman, 2004).

To address these two challenges, we construct a novel database of *patent interferences*. Patent interferences measure (nearly) *simultaneous* instances of *identical* invention by two or more *independent* parties. Under the "first to invent" rule prevailing in the U.S. until 2013, the U.S. Patent and Trademark Office (USPTO) investigated independent, simultaneous, and identical claims of invention in a patent interference. The inventor who conceived and reduced to practice first was awarded patent protection. (This system contrasts with the "first to file" rule more common in the rest of the world and prevailing in the U.S. today. Under "first to file," the inventor who files first wins the patent.) Famously, simultaneous claims by Alexander Graham Bell and Elisha Gray in February 1876 provoked an interference over the telephone patent.

By recording instances of common invention, patent interferences create a unique record of common knowledge inputs. Historians of science have noted the frequency of simultaneous, independent discovery, which Merton (1973) called "multiples"—e.g., the independent invention of calculus by Newton and Leibniz, or the independent formulation of the theory of natural selection by Darwin and Wallace. The view that new ideas result from combinations of existing ideas (Weitzman, 1998) suggests that interfering inventors share similar existing knowledge. For example, interfering inventors may have a similar background in chemistry, or they may have similar information about market conditions. Consider Gladwell's (2008) description of Bell's and Gray's invention of the telephone: "They arrived at electric speech by more or less the same pathway. They were trying to find a way to send more than one message at a time along a telegraph wire—which was then one of the central technological problems of the day. They had read the same essential sources—particularly the work of Phillipp Reis, the German physicist who had come startlingly close to building a working telephone back in the early eighteen-sixties."

We use interferences to test the hypothesis that geographically proximate inventors are more likely to share common knowledge inputs. The commonplace nature of multiples has led some to speculate that they must be "in the air, products of the intellectual climate of a specific time and *place*" (Gladwell, 2008; our emphasis). If knowledge spillovers are geographically localized, then a pair of neighboring inventors will have access to the same input knowledge. If they have access to the same input knowledge, they may be more likely to use it successfully in the same way, provoking a localized interference. However, if access to required knowledge inputs is unaffected by distance, then interference would be as likely between neighboring inventors as between distant inventors.

We show that interfering inventor pairs tend to be geographically concentrated, suggesting that neighboring inventors are more likely to share common knowledge inputs. Central to identification is whether or not the co-location of interfering inventors can be attributed to localized knowledge spillovers or some other factor. For example, inventors may co-locate to take advantage of thick markets for specialized skills (Bleakley and Lin, 2012), or they may co-locate to share a fixed, non-traded physical input (Helmers and Overman, 2017). To deal with the problem of inference under multiple sources of agglomeration economies, we use a matched control approach following Jaffe et al. (1993). For each interfering pair of inventors, we create a control pair by matching an interfering patent or application with a control patent matched on technology class and the date of invention. The control pair represents the expected proximity of inventors working in the same field and time period, except not conditioned on a "knowledge spillover" (i.e., an interference).

We find that interfering inventor pairs are 1.4 to 4.0 times more likely to live in the same local area—a place, city, or region—compared with control inventor pairs. Identification relies on (i) interfering inventors sharing common knowledge inputs and (ii) matching on observables fully accounting for other factors besides localized knowledge spillovers affecting the geography of invention. We show that our results are robust to conditioning on additional controls. Further, as Jaffe et al. (1993) note, to the extent that control pairs also co-locate to take advantage of knowledge spillovers, we will tend to under-estimate the importance of localized knowledge spillovers. In addition, we avoid scale and border problems by using distance-based tests of localization instead of aggregating inventors to administrative spatial units of arbitrary size (e.g., cities or counties) (Duranton and Overman, 2005). For example, measuring co-location at the county or commuting zone level understates localization if inventors are clustered on opposite sides of a county boundary. Murata et al. (2014) show downwards bias from scale and border problems is large relative to the upwards bias from imperfect matching as emphasized by Thompson (2006) and Thompson and Fox-Kean (2005).

Interfering inventor pairs are also more geographically concentrated compared with *citation-linked* inventors. Pairs of cited-citing inventors may provide an even closer match compared with matching only on other observables. The localization of interfering inventors compared with citation-linked inventors—who likely share codified knowledge—points to the important role of geographic proximity in facilitating flows of tacit, or hard to codify, knowledge. We also analyze the role of previous co-inventor ties in mediating the relationship between localization and interference. Inventors may be linked by a social network defined by inventors who have previously been listed as co-inventors on a patent. Inventor pairs linked by previous co-inventor ties are more likely to interfere with each other compared with inventor pairs not linked by previous co-inventor ties. However, in contrast to evidence from Breschi and Lissoni (2005) and Head, Li, and Minondo (forthcoming), we find little evidence that co-inventor ties are an important channel for the localization of interferences.

Our approach departs from existing ones that use the "paper trail" of patent *citations* to identify localized knowledge spillovers. In principle, a patent citation measures a piece of existing knowledge upon which the new patent builds. Earlier work has noted that citations are a "noisy" signal of knowledge spillovers. A survey of patenting inventors by Jaffe et al. (2000) reported that "one-half of all citations do not correspond to any spillover," and Jaffe et al. (1993) acknowledged that "an enormous number of spillovers [occur] with no citation."[1] Our contribution is to focus on identical inventive *output*. Thus, we circumvent the requirement of exactly measuring flows of knowledge across inventors using citations as measures of *inputs*.[2] Instead, our view is that identical invention likely requires a broad range of shared knowledge inputs. Citations may capture some of these shared knowledge inputs. But many knowledge inputs are not measured by citations—because they are not patentable and therefore not citable, or perhaps because they are tacit. Thus, interferences may better capture spillovers of tacit knowledge, or at least knowledge that is not measured by a patent citation.[3]

Second, a common interpretation of a citation is that knowledge has "spilled over" from the cited inventor to the citing inventor. Our setting is more agnostic: it doesn't matter where input knowledge originates. It might have been discovered by an interfering inventor, or it may have been discovered by a friend or colleague (or a more passing connection) of an interfering inventor. The key idea is that the accessibility of knowledge varies across space, so that co-located inventors are more likely compared with distant inventors to share the same knowledge.[4]

1 Identifying localized knowledge spillovers with interferences

We test for localized knowledge spillovers by comparing the geographic proximity of interfering inventor pairs with a set of matched control pairs, following the strategy of Jaffe et al. (1993). An interfering inventor pair consists of two independent inventing teams declared in interference by the USPTO, having claims covering "the same, or substantially the same, subject matter" (35 U.S. Code §135). A matched control pair includes one of the interfering applications and one

[1]Inventors strategically not citing known prior art contribute to this problem.

[2]Further, compared with earlier efforts to measure multiples (Ogburn and Thomas, 1922; Bikard, 2012), our database is large and relies on the real-time declaration of multiples by patent examiners instead of *ex post* measurement by researchers. In a contemporaneous paper, Baruffaldi and Raffo (2017) use citations classified by European Patent Office examiners as showing the claimed invention is not novel to construct a database of duplicated inventions. In contrast, interferences seem likely to measure duplication beyond cited prior art.

[3]Conversely, like citations, interferences are an indirect measure of shared knowledge inputs. Our contribution is that interferences seem likely to be generated in a very different manner compared with citations, providing important complementary evidence on the localization of knowledge spillovers.

[4]One could also re-interpret the meaning of patent citations in a similar manner: Instead of a direct spillover from the cited to the citing inventor, a citation is more likely when the citing inventor is more likely to learn about the cited patent from local ties.

control issued patent matched on technology class and application date (within 180 days of the application date of the interfering application.) The idea is that inventors of the control patent likely face a similar location choice problem compared with interfering inventors. By comparing the geographical localization of interfering inventors to this counterfactual, we hope to control for all factors except common knowledge inputs.[5]

Consider the following reduced-form model describing the relationship between shared knowledge, location, and interference. The probability of interference $P(int_i)$ between a pair of inventors i depends on their shared knowledge inputs A_i, observable factors X_i, and an unobservable factor n_i,

$$P(int_i) = g(A_i, X_i, n_i). \tag{1}$$

The degree of knowledge inputs shared by an inventor pair is a function of the geographic distance between them, $dist_i$, and an unobservable factor e_i,

$$A_i = f(dist_i, e_i). \tag{2}$$

We compare the proximity of interfering pairs with that of control pairs. This will identify localized knowledge spillovers under two assumptions. First, interferences must be more likely to be declared between patents whose inventors have more knowledge inputs in common—that is, g must be monotonically increasing in A_i. This assumption follows from the logic that new ideas result from combinations of existing ideas, and therefore inventors who make the same discovery must share the same knowledge inputs. Second, the matched control pairs must fully account for other factors *besides localized knowledge spillovers* affecting the geography of interference. That is, n_i must be independent of $dist_i$. Under these assumptions, a positive correlation between interference and proximity (conditioned on observed factors X_i) is evidence of localized knowledge spillovers.

First, we use non-parametric, distance-based methods (Duranton and Overman, 2005; Murata et al., 2014) to test for localized knowledge spillovers. We compute distances between interfering inventor pairs to overcome scale and border problems. Then, we compare the distribution of the distances between interfering inventor pairs to a counterfactual distribution of distances between control inventor pairs. We construct confidence intervals to test for significance differences between the localization of interfering and control pairs. This analysis is described in Section 3.

Second, we use a linear probability model which compares the probability of interference between patent pairs above and below different co-location thresholds. The linear probability models we estimate can be motivated by assuming that equations (1) and (2) are linear in their arguments.

[5] As Jaffe et al. (1993) note, this may be a conservative estimate in the sense that only the concentration *in excess* of the control pairs is attributed to localized knowledge spillovers. However, if matched control pairs do not adequately control for unobserved factors (other than knowledge spillovers) affecting the geography of invention then our estimate of knowledge spillovers will be biased upwards.

This analysis is described in Section 4.

A central concern is that control pairs of patents may imperfectly capture unobservable factors. This is the subject of the citation analysis by Thompson and Fox-Kean (2005), who show that geographic matching results following Jaffe et al. (1993) are sensitive to matching on technological classification. In particular, conditioning matched control patents on (6-digit) technology *subclass*, the finest detail available in the U.S. patent classification, rather than the 3-digit technology class in Jaffe et al. (1993), shows little localization of citations compared with control pairs.[6]

A useful set of results comes from Murata et al. (2014). First, the degree of localization of citations appears to be considerably understated by the geographic matching tests of Jaffe et al. (1993) and Thompson and Fox-Kean (2005). Murata et al. (2014) show that a *distance-based* test shows localization of patent citations, even when selecting matched controls of 6-digit subclasses. The reason that geographic-matching tests understate the localization of citations is that aggregating inventor locations to metropolitan areas and states introduces border and scale problems. Border problems arise because counties, the basis of metropolitan areas, may split clusters of inventors. This tends to bias downwards measures of localization, since inventors split by a county boundary will not be localized by a commuting zone measure. Scale problems arise because aggregating the data to metropolitan areas or commuting zones allows only analysis at one spatial scale. Since commuting zones vary from very small to very large in area, analyses based on commuting zones mix different spatial scales. Thus, the sensitivity of the results on the localization of citations appear to be dwarfed by the (opposite) bias introduced by spatial aggregation. Murata et al. (2014) perform a sensitivity analysis following Rosenbaum (2002) to bound the degree of bias in the presence of unobserved factors affecting localization. They find that biases from such unobserved factors would have to be extremely large before reversing the conclusion that patent citations are geographically localized.

Aside from concerns about imperfect controls, one might also be concerned that as inventions become more technologically similar (in "idea space"), that other sources of agglomeration economies—say, labor pooling—might also become stronger. If this were true, even a "perfect" control patent would be unable to distinguish localized knowledge spillovers from other factors.[7] This is because the "idea similarity" between interfering inventors reflects not only shared knowledge inputs but also greater similarity in other inputs. For example, if there are more potential employees who know about the details of reprogramming yeast cell reproduction in San Francisco, then inventors who want to reprogram yeast cells and are concerned about insuring against the idiosyncratic risk of losing an employee will be more likely to locate in San Francisco.

[6] Henderson et al. (2005) note considerable uncertainty in the "right" way to select matched controls, whether at the 3-digit class level (with about 450 classifications) or the 6-digit subclass level (with about 150,000 subclasses). While 6-digit technology classifications are quite detailed (Thompson and Fox-Kean (2005) note that the 3-digit class "231–Whips and whip apparatus" contains 7 distinct subclasses), Henderson et al. (2005) counter that there is little evidence suggesting that selecting on 6-digit subclasses achieves "closer" technologically matched controls.

[7] We are grateful to an anonymous referee for highlighting this issue.

On the other hand, localized knowledge spillovers may still be identified using a matched-control approach if the "transportation cost" of ideas exceeds that of other factors. Suppose the strength of knowledge spillovers declines swiftly with distance—chance meetings or casual conversations with social ties are much less likely even at modest distances (Allen, 1984; Arzaghi and Henderson, 2008)—but labor pooling benefits are realized across neighboring counties within the same commuting zone. In other words, the transportation cost of ideas—particularly tacit knowledge—is higher compared with the cost of transporting workers, i.e., commuting, or the cost of transporting other factors (Rosenthal and Strange, 2001). Firms that benefit from *both* localized knowledge spillovers and labor pooling will be more likely to co-locate in the same office park or neighborhood compared with two firms that benefit *only* from labor pooling, but not from localized knowledge spillovers.

2 Data

2.1 Patent interferences: Background

Patent interferences were a unique feature of U.S. patent law. Through March 16, 2013, the U.S. had a "first to invent" rule for assigning priority of invention, versus the "first to file" rule more common in the rest of the world and that prevails in the U.S. today. When the USPTO received patent applications from multiple, independent parties with one or more identical claims at roughly the same time, it was obliged to investigate and determine which party was entitled to patent protection. This investigation was known as a patent interference. A rotating three-judge panel from the Board of Patent Appeals and Interferences (hereafter Board) decided who was first to invent, meaning both (i) who first had the idea (conception) and (ii) who first put the idea in workable form (reduction to practice). Typically, parties submitted dated laboratory notebooks, testimony by associates, and media reports as evidence of first invention.[8] Many interferences involved valuable inventions.[9]

Interferences are distinct from patent infringements. First, an interference was suggested by a patent examiner, a specialist in a particular technological area, during their routine search for prior art, when at least two U.S. patent applications (or one patent application and a recently-issued patent) contained the "same patentable invention" (37 Code of Federal Regulations §1.601).[10] The

[8]More details about the patent interference proceedings can be found in Calvert (1980), Calvert and Sofocleous (1982), Cohen and Ishii (2006), de Simone, Gambrell and Gareau (1963), and Kingston (2001). Lin (2014) reviews this literature and provides summary statistics for the patent interferences used in this study.

[9]A recent interference case decided in February 2017 involved patent rights to the CRISPR gene-editing technique. (The decision date puts the case outside our sample.) In that decision, the Board found no interference in fact—the inventions claimed by the competing inventors, assigned to the Broad Institute and the University of California, were separate and did not overlap. The validity of the Broad patents was a "surprise" to researchers in the field; as a sign of the value of the invention, the stock of the licensee to the Broad patents went up sharply following the decision (Pollack, 2017).

[10]Alternatively, an examiner could suggest an interference when "the subject matter of a claim of one party would,

patent examiner would then forward the patent application and a memorandum to the Board, which would declare the patent interference. This is distinct from a patent infringement, in which the holder of an existing patent sues an infringing party.[11] Second, interferences must involve parties with nearly simultaneous pending applications for patents.[12] This feature makes interferences distinct from patent infringements, which typically involve leaders and followers.

2.2 Database of interfering inventor pairs

We constructed a database of patent interference cases for our analysis. Our database starts with information from 1,329 interference *decisions* issued by the Board between 1998 and 2014.[13] These decisions were downloaded from the Board's "e-FOIA Reading Room." From each decision, we record information about the *case*, the *parties*, the *application(s)* and/or *patent(s)*, the *claims*, and the *inventors*.[14] A typical decision reports the names of the interfering inventors and assignees, the associated patent and application numbers, and the decision on priority at the claim level or other disposition of the case. We also collect additional details on the cases and inventors from the USPTO's "eFile" site or the Patent Application Information Retrieval (PAIR) service. The eFile site sometimes contains notices of settlement agreements.[15] See Appendix A for more details.

Several features of interference practice allow us to rule out alternative explanations, outside of shared knowledge inputs. First, interferences between parties with common ownership interests were not allowed. Thus, they seem unlikely to result from other shared factors or from within-firm spillovers. Second, the Board might decide that there was no interference in fact. This

if prior art, have anticipated or rendered obvious the subject matter of a claim of the opposing party and vice versa" (37 C.F.R. §41.203).

[11]In some cases, a patent applicant could suggest a possible interference, but private parties could not sue for an interference.

[12]The claim(s) of invention must have satisfied standard patentability rules—i.e., the claims must have been in an patent-eligible class, useful, novel, and non-obvious. In addition, the USPTO required that a timing rule be satisfied in order to avoid interferences resulting from the disclosure of patent applications themselves (i.e., publicized patent applications leading to copycat inventions). Thus, in the case of two or more interfering applications, the dates of application must have been no more than 3 months apart. In the case of an interfering issued patent and pending application, (a) the application's date must have been more than one year before the patent's grant date and (b) the application's date must have been no less than 3 months after the patent's application date. Specifically, interfering claims among pending applications must be made within 1 year of each other (35 U.S.C. 135.b.2). In cases where an application's claims interfere with an already-issued patent, the claims must be made no later than 1 year prior to the patent's issue date (35 U.S.C. 135.b.1), and typically no later than 3 months after the patent's original application date (37 C.F.R. 1.608).

[13]There are a few decisions related to interferences declared well before 1998, as far back as the early 1980s, including a famous case over the method of producing the hepatitis B antigen. On average, however, the lag between interference declaration and decision dates is a few years. See Calvert and Sofocleous (1982, 1986, 1989, 1992, and 1995.)

[14]Note the following: (1) each case is argued between two or more parties; (2) each party may have one or more inventors; (3) each party may also have one or more applications and/or patents in interference; (4) each application or patent makes one or more claims; (5) one or more of these claims are declared by the examiner to be in interference.

[15]Normally, details of these agreements were kept secret. These notices acknowledge the existence of a settlement agreement, as opposed to a decision on priority or some other outcome. Unfortunately, settlement agreements are sealed. Thus, we can note their existence, but we cannot analyze their contents.

helps us to distinguish between identical inventions and near misses.[16] Third, interferences seem unlikely to result from intentional delay in disclosing an invention. According to interference rules, an interfering inventor would lose priority if the inventor had not immediately filed for a patent application following conception and reduction to practice.[17] The disclosure of the timeline of invention was the primary purpose of the interference proceeding. The alternative to admitting an intentional delay in applying for a patent would be to concede a later date for reduction to practice, weakening the case for priority.[18] Fourth, if one party's application was derived—i.e., that one inventor's claims are directly sourced from the competing party through stealing or espionage—that was grounds for an adverse judgment.[19]

Interference cases were terminated by the judges' decision on priority or for some other reason. A decision on priority meant a judgment that one party had first conceived of the invention and reduced it to practice. Table 1 summarizes case dispositions for our sample. The first two columns display frequencies and the share of cases by disposition for our full sample. Nearly 20 percent of cases resulted in a judgment on priority, while nearly 60 percent of cases were conceded. Concessions occur when one party files a request for adverse judgment. An abandonment occurs when one party fails to file at some stage of the case. We code these outcomes as they are noted in the decisions. Absent detail in the decision, it is difficult to ascertain the motivations for concessions (and we cannot rule out abandonments if failure to file is not mentioned.) However, for the sub-sample of 981 cases that we are able to match to documents on the eFile site, we code cases including an acknowledgement of settlement. In this sub-sample, settlements constitute the majority of concessions and one-third of all cases. We are unable to characterize the remaining conceded cases that have no acknowledgement of settlement or text in the decisions referring to a failure to file.

[16]Interferences appear to involve similar inventive claims, based on the counts declared by the Board and the corresponding claims of invention in interfering applications. In Appendix B, we show that interfering applications overlap substantially in their inventive claims. On average, over three-quarters of a party's application claims are found to be in interference. Further, partial decision are rare. Typically, the interfering claims are awarded entirely to one party or the other. Thus, *contra* Schmookler (1966), interferences can identify identical inventions, versus near misses.

[17]Cohen and Ishii (2006) argue that interferences correspond to an incumbent-entrant game where incumbents decide to keep inventions secret for some period of time before filing a patent application. The requirement to promptly disclose an invention may have reduced the relevance of this margin as decisions went against interfering inventors who chose to keep their inventions secret for some time.

[18]Interference cases appear to vary in terms of whether inventors are aware of each other's efforts. In Lutzker v. Plet (1988), the United States Court of Appeals, Federal Circuit, affirmed that Lutzker was not entitled to a patent for a canape maker, despite having established conception and reduction to practice in early 1976, since he had delayed disclosure and filing for a patent until late 1980. In contrast, Plet received priority by demonstrating her conception and reduction to practice by early 1980, with a filing date of March 3, 1980. The original decision by the Board cited the failure of Lutzker to show renewed activity towards disclosure "until after Plet entered the field" as an important factor in the judgment against him.

[19]Cases involving stealing or espionage could still reflect a spillover if it is a knowledge input rather than the complete invention that was stolen. The "derivation" judgment (see Table 1) occurred when it could be proved that one party completely stole the invention from the other. This judgment is very rare and so including them in the analysis does not affect our results – there is only one case with a "derivation" outcome out of the 1,329 in our database.

Table 1: Sample distribution of interference case dispositions

Disposition	*Full sample*		*eFile*
Number of cases	1,329		981
Decision on priority	260	19.6%	19.0%
Conceded, total	781	58.8	57.1
... settled	.	.	34.2
... abandoned	92	6.9	3.5
... all other reasons	.	.	19.4
No interference in fact	46	3.5	3.4
Common ownership	64	4.8	4.8
Unpatentable	122	9.2	11.3
Other	56	4.2	4.6

Most of our analysis focuses on interference cases where the board's decisions report a settlement or judgment on priority. Shared knowledge inputs seem more likely in these cases compared with other case dispositions. For example, cases dismissed for no interference in fact seem less likely to involve exactly common knowledge inputs. Other dispositions were less common. About 9 percent of cases were dismissed because the claims were deemed unpatentable. Five percent were dismissed because the interfering parties were discovered to have assigned rights to a common owner, e.g., a multinational firm.[20] Three percent of cases were dismissed after a finding of no interference in fact.

2.3 Database of control inventor pairs

We use USPTO technology classifications and application dates to select controls. We obtain a list of all the technology classifications assigned to a patent from the USPTO's Master Classification File. Patents are classified according to 3-digit technology classes and 6-digit subclasses. This information is available for all issued patents, but only for patent applications filed in 2001 and later. For earlier patent applications, we instead obtain classification information from the PAIR service, which records only a single main technology classification, considered the "primary classification" of that invention.

Since some interference cases involve more than two parties, our database of 1,329 interference cases involves 1,401 interfering pairs of inventing parties.[21] We select a set of control patents that are similar to the invention described by each party's patent(s) or application(s) declared in interference. First, we require a control patent to share at least one 3- or 6-digit technology

[20] In at least one case, a merger appeared to have been *caused* by the pending interference.

[21] Twenty-four cases involve three parties, three cases involve four parties, and one case involves five independent parties.

classification with a party's patent(s) or application(s). Second, control patents must have an application date within 180 days of the application date of an interfering application. We are able to obtain a set of suitable control patents for nearly every interfering inventor pair—only 24 pairs lack suitable 3-digit controls and 32 pairs lack suitable 6-digit controls.

Next, each of these control patents is now eligible to form a *control pair*. A control pair matches a control patent to an interfering patent or application. Control patents are matched to the opposing party's interfering application(s). This matching structure is identical to Jaffe et al. (1993). In their application, *cited* patents are used to identify control patents. Control patents are then matched to *citing* patents. In our sample, the interfering application(s) of the first party are used to identify control patents. Control patents are then matched to the interfering application(s) of the second party. The pool of control patents is large. Conditioned on finding a control, the average interfering inventor pair is associated with 706 control patents matched on 6-digit technology class and 6,457 controls matched on 3-digit technology class.

2.4 Inventor location

We use information from patents and applications to measure the locations of inventors. For issued patents, we use the inventor disambiguation dataset of Lai et al. (2013). If an interference is decided against a party without an issued patent, then the losing party's patent application is never passed for issue. In this case, the decisions, the notices of interference from eFile, and the PAIR database are essential sources of inventor location. While 75 percent of inventors are located in the Lai et al. (2013) database, the remaining 25 percent of inventors' locations are recorded *only* in the decisions, eFile, or PAIR. At the case level, we are able to record inventor locations for at least 1 inventor in each party in 88 percent of interference cases, and all inventors in 85 percent of interference cases.

We compute the distance between opposing pairs of interfering or control inventors. We use the place of residence of inventors—the place named in the bibliographic description of the patent or application. Places can be large cities (e.g., San Francisco, CA), but also can be small towns (e.g., Burlingame, CA) or even unincorporated places. Unfortunately, the data are not detailed enough to describe co-location within a place. We obtain the latitude and longitude of each inventor's place of residence from the Census Gazetteer file. We then compute the minimum geodesic distance across all possible pairings of inventors within an interfering or control pair.[22]

We also construct a geographic matching measure of co-location, in order to provide results comparable to earlier work and to highlight proximity at short distances. For example, we define

[22]For an interference between a party with m co-inventors and another party with n co-inventors, there are mn pairwise combinations of inventors. We report results using the minimum pairwise distance across all of these pairwise combinations. We also experimented with using the median or mean distance for each interfering pair of parties; as well as, following the convention of earlier work, the distance between the first-named inventors of each patent. Results using alternative measures are similar to those reported in the paper.

a variable that indicates whether or not an inventor pair shares the same place of residence, or if their places of residence are within within 161 kilometers (100 miles) of each other. By these measures we intend to capture localized interactions via social ties, workplace relationships, or random meetings. The 100-mile cutoff is comparable to a metropolitan area (used in Jaffe et al., 1993) or a commuting zone (as in Autor and Dorn, 2013).[23] An advantage of our distance-based measure compared with commuting zones is they avoid border and scale problems as described in Section 1.

2.5 Illustrative example

We briefly discuss one interference case to illustrate how interferences reflect shared knowledge inputs and how the matched control strategy works in practice. Interference 103,435 involved competing claims of invention of an intraocular lens—a lens that is implanted in the eye during procedures such as cataract removal. The claims at issue involved the method by which haptics, side struts which hold the implanted lens in place inside the eye, are attached to the optic (lens). Both parties found that exposing the haptics to corona discharge (a plasma curtain that is created when air around a conductor gets ionized) increases the strength of the bond between the haptic and the lens.

The junior party included 3 employees (Richard Christ, David Fencil and Patricia Knight) in the R&D Department at Allergan Inc., a pharmaceutical company located in Irvine, California.[24] The senior party was Larry Blake, inventor and owner of the small company Pharmacia Iovision, Inc., which was also located in Irvine. The Board determined that Blake had both conceived the invention and reduced it to practice sometime in August 1987, while the Christ team had conceived the invention sometime in 1985 and reduced the invention to practice in December 1987.[25]

A few aspects of the case are relevant for our analysis. These details are described in the Board's decision, which was based on evidence presented during the case, primarily lab notebooks and eyewitness testimony.

First, both parties appear to have had common knowledge inputs that led to the common discovery. They were aware of the problem of attaching haptics to lenses and knew that corona discharge had been "used for surface treatment of plastics and rubber to improve adhesion."

Second, the geographic proximity of the inventors facilitated knowledge flows. Most relevantly, Blake had been a short-term consultant at Allergen between June 1986 and April 1987. His consulting work at Allergan related to silicone lenses. Notably, he was *not* working with the Christ team nor in a similar research area. Nonetheless, the decision cites testimony from Allergen employees

[23] On average, commuting zones are similar in size to a circle with diameter 80 miles; the largest consolidated statistical area (New York) is similar in size to a circle with diameter 130 miles.

[24] Allergan initially had an R&D focus on eye care therapies and its flagship product is Botox. It purchased American Medical Optics, located in Irvine in 1986. It was one of the largest intraocular lens producers at the time.

[25] Priority was awarded to Blake as Christ "failed to show diligence between the period just prior to August 1987 and the reduction to practice date established for the Junior party of December 31, 1987."

stating that Larry Blake would have been aware of the project through his physical presence at Allergen: "...there were no secrets in the R&D Department of Allergan during the mid-1986 time frame and everyone shared resources." One employee testified that the "plasma and corona work was being conducted in the Technology and Ventures laboratory" and Blake was seen at least once in the building and "anyone could go in and out of the building freely."

Both firms were located in Irvine, and the individual inventors lived nearby: Blake, Christ, and Knight lived in surrounding Orange County communities (in Cota de Caza, Laguna Beach, Laguna Niguel, respectively), while Fencil lived in Goleta, in Santa Barbara County. (These places of residence are the locations recorded in their corresponding patents and applications.)

Third, examining the control patents matched to the inventions in interference suggests that the control inventions are quite similar to those in the dispute, but the inventors are geographically distant. For example, one control patent matched to the Christ application was filed by 3 inventors all living in the Minneapolis–St. Paul metropolitan area and working at Minnesota Mining and Manufacturing Company (3M) in St. Paul, Minnesota. The invention appears quite similar to the interfering patents; it is for an intraocular lens and discloses a method for fixing a haptic having an anchoring filament to a lens. Another control patent matched to the Blake application was filed by an inventor living in Arlington, Texas, and also concerned attaching a haptic to a lens for an intraocular lens (titled "Method of Attaching a Haptic to an Optic of an Intraocular Lens").

3 Distance-based and geographic matching tests of localization

We identify localized knowledge spillovers using distance-based tests following Duranton and Overman (2005) and Murata et al. (2014). We compare the distribution of distances between pairs of interfering inventors to a counterfactual distribution of distances between pairs of control inventors. In this way, we compare the geographic distribution of interferences with the distribution that would be expected to occur randomly conditioned on the geography of invention represented by the controls.

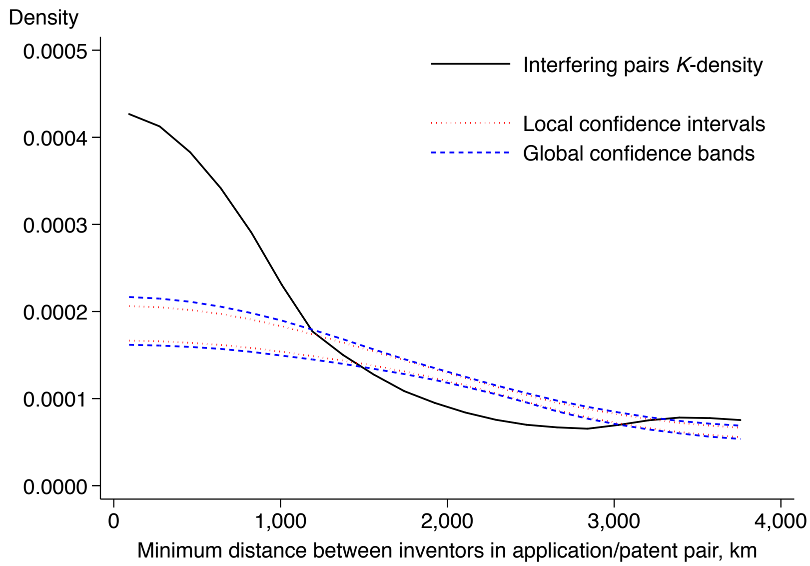
First, we estimate the density of distances between pairs of interfering inventors. The distance between each interfering inventor pair $i = 1 \dots I$ is d_i. The estimator of the density of pairwise distances (the kernel density) at any distance d is

$$\hat{K}(d) = \frac{1}{2h} \sum_{i=1}^{I} f\left(\frac{d - d_i}{h}\right),$$

where h is the bandwidth and f is the kernel function. All densities are computed using a Gaussian kernel and with the bandwidth set as in Silverman (1986).[26] Figure 1 shows the estimated density

[26]To deal with boundary problems at zero due to the non-negative domain of distances, we use the reflection method of Silverman (1986), following Duranton and Overman (2005).

A. Interfering and 3-digit control pairs



B. Interfering and 6-digit control pairs

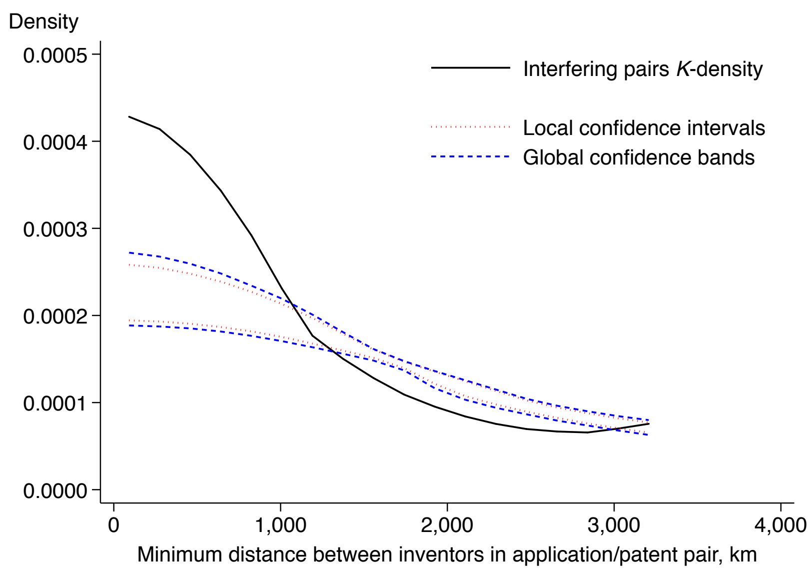


Figure 1: Interfering pairs are more localized compared with control pairs

These graphs compare the estimated kernel densities of geographic distances between pairs of interfering inventors to the distribution for similar non-interfering control pairs. We use the minimum geodesic distance between the places of residence of inventors from opposite teams. The black line shows the estimated kernel density function for all interfering pairs for which we were able to find suitable controls. The dotted red lines show the upper and lower 5% local confidence intervals of the kernel density for non-interfering control pairs. The dotted blue lines show the upper and lower 5% global confidence bands of the kernel density for non-interfering control pairs. The estimated densities end at the median distance between inventing pairs. Interfering pairs are considered localized if the estimated kernel density of interfering inventors is above the upper global confidence band for at least one distance d up to the sample median.

of pairwise distances between interfering inventors as a solid black line.

Second, we use Monte Carlo simulations to construct a counterfactual distribution of pairwise distances using our sample of control pairs, test for significant departures from our counterfactual benchmarks, and estimate confidence intervals. Our simulation strategy follows Murata et al. (2014). Our analysis differs from Murata et al. (2014) by focusing on pairs of patents in an interference case, rather than cited-citing pairs.

In each of 1,000 simulations, we randomly sample a matched control pair for each interfering pair from the set of permissible control pairs. With this random draw of control pairs, we estimate the kernel density for the distribution of pairwise distance. We repeat this exercise by drawing a new set of control pairs.

We evaluate significant departures from our counterfactual benchmarks by constructing local confidence intervals. We consider all distances between 0 and 3,844 km, the median distance between all pairs of control and interfering inventors.[27] Comparison of the distance densities above 3,844 km is redundant because densities must sum to one over the entire range of distances. In other words, if the density of interfering pairs is *lower* compared with control pairs at distances greater than 3,844 km, it must be *higher* compared with control pairs at distances less than 3,844 km. Thus, we need only examine below-median distances to infer localization.

We rank the simulated kernel density estimates at 100 evenly spaced distances, and select the 50th ranked simulated kernel density at each distance to construct the lower 5% confidence band and the 950th-ranked to construct the upper 5% confidence band. These are the lower 5% and upper 5% confidence levels denoted by $\underline{K}(d)$ and $\overline{K}(d)$, respectively. When $\hat{K}(d) > \overline{K}(d)$, we infer that interferences exhibit localization at distance d at a 5% confidence level. That is, interfering pairs are more likely to be d km apart compared with control pairs. Figure 1 shows that the estimated kernel density of pairwise distances for interfering pairs exceeds the upper 5% local confidence level (dotted red line) for a range of distances up to about 1,000 km (620 miles). For example, since $\hat{K}(d) > \overline{K}(d)$ for $d = 1000$, we infer that interfering inventor pairs are more likely (at a 5% confidence level) to be 1,000 km apart compared with control inventor pairs.

These inferences are *local* in the sense that they only allow us to make local statements (i.e., at a given distance d) about the relationship between the interfering and counterfactual distributions. However, even if interfering pairs were distributed randomly with respect to control pairs, there is a high probability that interfering pairs will exhibit localization at some distance, since by construction there is a 5 percent probability for each particular distance that a random draw of control pairs will exhibit localization.

We next define global confidence bands, which allow us to make statements about the overall location patterns of interfering inventors. We search for identical upper and lower confidence intervals such that when we consider them for all distances between 0 and 3,844 km, only 5% of

[27] The median distance is 3,367 km for the sample of interfering and 6-digit control pairs and 3,844 km for the sample of interfering and 3-digit control pairs.

our randomly generated simulated kernel densities hit them. That is, we define a global upper confidence band $\overline{\overline{K}}(d)$ as the band that is hit by 5% of our simulations between 0 and 3,844 km. Interfering pairs are considered globally localized (at a 5% confidence level) when $\hat{K}(d) > \overline{\overline{K}}(d)$ for *at least* one $d \in [0, 3844]$. Naturally, the global confidence bands are wider compared with the local confidence intervals.[28] Graphically, interferences are globally localized if the density of pairwise distances for interfering pairs is above the upper global confidence band for at least one distance d up to the sample median. Conversely, interferences are globally dispersed if the density lies below the lower confidence band and never lies above the upper confidence band. Figure 1 shows that the estimated kernel density of pairwise distances for interfering pairs exceeds the upper 5% global confidence level (dashed blue line). Thus, interfering pairs of inventors are more geographically localized compared with control inventor pairs not linked by an interference.

These results are robust to matching on 3-digit technology class or 6-digit technology subclass. In panel A, we show results using 3-digit control pairs; in panel B, we show results using 6-digit control pairs. Thus, while the observed distribution of pairwise distances for interfering pairs is the same in both panels, the counterfactual distribution is more geographically concentrated for the 6-digit control pairs shown in panel B. This is consistent with Thompson and Fox-Kean (2005), who show that the localization of patent citations is sensitive to the selection of controls. However, unlike Thompson and Fox-Kean (2005), we find that for either counterfactual, interferences are indeed significantly geographically localized. This echoes Murata et al. (2014), who show that the localization of patent citations is robust to matching on 6-digit subclasses when using a distance-based test, as we are doing here. Overall, this evidence is consistent with geographic proximity facilitating the sharing of common knowledge inputs.

Our results on geographic localization are robust to the choice of proximity measure and conditioning on decision type. Table 2 presents results separately for priority decisions (excluding concessions) and three different measures of inventor proximity. Panel A compares the average distance between inventors in an interfering pair, to that of control pairs, with simulated confidence intervals. On average, interfering pairs in priority decisions and concessions are 3,451 km apart, compared with 4,778 km separating 3-digit control pairs of inventors and 4,425 km separating 6-digit control pairs of inventors. Average distances separating interfering and control pairs of inventors are similar when we focus on priority decisions only.

Though interfering inventors are closer together on average, the average pairwise distances may obscure the relevant range of geographic proximity for localized knowledge spillovers. Panels B and C present results that examine inventor co-location or "geographic matching" as in the main tests performed by Jaffe et al. (1993). Panel B shows the share of inventor pairs that report the same place (e.g., a town or city) of residence. Nearly 3 percent of interfering inventors share a

[28] In general, we end up selecting approximately the 10th-ranked simulated kernel density to construct the lower 5% global confidence level and the 990th-ranked simulated kernel density to construct the upper 5% global confidence level. See Duranton and Overman (2005) for more discussion.

Table 2: Interfering pairs are co-located compared with control pairs

	Type of case	
	Priority decisions and concessions	Priority decisions only
A. Average distance between pair of inventors in kilometers		
Interfering pairs	3,451	3,603
3-digit control pairs	4,778	4,714
	(4,563, 4,985)	(4,321, 5,122)
6-digit control pairs	4,425	4,282
	(4,219, 4,623)	(3,890, 4,688)
B. Share of inventor pairs with same place, town or city of residence		
Interfering pairs	2.7%	2.8%
3-digit control pairs	0.8%	0.7%
	(0.4%, 1.4%)	(0.0%, 1.8%)
6-digit control pairs	2.0%	2.0%
	(1.3%, 2.9%)	(0.7%, 3.6%)
C. Share of inventor pairs with places of residence within 161km or 100mi		
Interfering pairs	13.8%	11.7%
3-digit control pairs	5.2%	5.2%
	(4.0%, 6.4%)	(3.1%, 7.6%)
6-digit control pairs	8.2%	8.2%
	(6.7%, 9.7%)	(5.4%, 11.2%)

This table reports statistics for interfering and control pairs of inventing teams. Panel A reports the average minimum distance between places of residences of inventing teams. Panel B reports the share of inventor pairs that share a place of residence. Panel C reports the share of inventor pairs where the minimum distance between places of residence is within 161 km. Simulated upper and lower 5% confidence intervals for control pairs shown in parentheses.

place of residence, compared with 1 and 2 percent of 3- and 6-digit control pairs, respectively. (The difference compared with 3-digit control pairs is statistically significant.)

Panel C shows the share of inventor pairs that report places of residence within 161 kilometers or 100 miles of each other. By this measure we intend to capture localized interactions via social ties, workplace relationships, or random meetings. The 100-mile cutoff is comparable to a metropolitan area (used in Jaffe et al., 1993) or a commuting zone (as in Autor and Dorn, 2013).[29] As described in Section 1, we prefer to use this distance-based cutoff as opposed to explicitly calculating commuting zone matches, because it avoids border and scale problems. Thus, this test is similar to the original matching-rate tests at the metropolitan area level reported by Jaffe et al. (1993), except that we leverage the micro-geography of inventor location compared with co-locating within the same county or set of counties.

Between 12 and 14 percent of interfering inventor pairs are within 161 km (100 miles), compared with 5 to 8 percent of control inventor pairs. These differences are significant. In sum, interfering inventor pairs are 1.4 to 4.0 times more likely to locate in the same place or region compared with control inventor pairs.

In Appendix C, we show additional results comparing interfering inventor pairs to control pairs. We provide evidence of shared knowledge inputs among interfering and control pairs. Interfering pairs are more likely to share technology classifications and backwards citations. These results connect interferences to more familiar measures of shared knowledge. They also motivate our next analysis conditioning on additional controls.

4 Interference-rate regressions

The result that interfering inventors are more localized compared with control inventors is also robust to conditioning on bibliometric measures of pairwise similarity. To show this, we assume that equations (1) and (2) are linear in their arguments and estimate:[30]

$$Pr(int_{i[g]}) = \mu_g + \beta_1 f(distance_i) + \mathbf{X}_i \beta_X + \varepsilon_i. \tag{3}$$

The dependent variable is a dummy indicating whether an inventor pair is involved in an interference. The regressions are run on the entire sample of interfering pairs and matched controls. All

[29] On average, commuting zones are similar in size to a circle with diameter 80 miles; the largest consolidated statistical area (New York) is similar in size to a circle with diameter 130 miles.

[30] This equation illustrates that the effect of distance on interference based on a naïve comparison between interfering pairs and arbitrarily-chosen patent pairs will be biased if omitted factor η_i that affects the probability of interference is correlated with the distance between the pair of inventors. Ideally, the matched-control estimator can still identify β_1, the effect of distance on interference. To see this, take expectations and note that by assumption $E(\epsilon|IC) = 0$. That is, conditioned on the sample of interfering and control pairs IC, the expected effect of unobserved factors on interference is zero. However, to the extent that matched controls do *not* satisfy this condition, then estimates of $\hat{\beta}_1$ will still be biased. In other words, perhaps the matched controls are "not similar enough," and therefore, the identifying assumption is not satisfied.

specifications include a fixed effect for each "pair group" indexed by g, defined by an interfering inventor pair and all associated matched control pairs. Thus, the effect of co-location on interference is identified by variation within the group of an interfering pair and its associated control pairs. We cluster standard errors at the pair-group level.

We use several measures of the proximity of inventor pairs, following the results reported in Table 2: (i) the logarithm of distance between pairs of inventor teams; (ii) an indicator equal to 1 when the minimum distance between an inventor pair is within 1.6 km (1 mile); and (iii) an indicator equal to 1 when the minimum distance between the locations of residence reported by an interfering or control pair is within 161 km (100 miles).

Estimates are reported in Table 3. For each combination of sample (interfering pairs plus 3- or 6-digit control pairs) and proximity measure, we show two specifications. The first, in columns (1) and (3), reports the univariate relationship between inventor pair distance and interference. The second, in columns (2) and (4), reports conditional estimates, controlling for three measures of bibliometric similarity: (i) the number of technological classes shared by the pair of inventors, (ii) the number of shared subclasses by the pair of inventors, and (iii) the number of backwards citations to prior art shared by the pair of inventors. Recall that we only match controls based on a single shared (3-digit) class or (6-digit) subclass. Including controls for additional shared classes and subclasses conditions our estimate of the effect of distance on even more similar inventions—that is, those pairs that share similar numbers of classifications. Similarly, including controls for the number of backwards citations conditions our estimates on pairs of inventions that probably, by evidence of citing the same prior art, are technologically quite similar.

Proximity leads to a higher likelihood of interference, as suggested by the estimates in Panel A. In the sample of 3-digit matched control pairs and interferences, a doubling of the minimum distance between a pair of inventor teams decreases the probability of interference by 0.008 percent (column 1), compared with the mean probability of interference of 0.04 percent. In the sample of 6-digit matched control pairs and interferences, a doubling of the minimum distance between a pair of inventor teams decreases the probability of interference by 0.04 percent (column 3), compared with the mean probability of interference of 0.4 percent. Similarly, in panels B and C, co-location in the same place of residence or with places of residence within 161 km increases the likelihood of interference by 0.06 percent and 0.04 percent, respectively, compared with 3-digit control pairs. These effects are precisely estimated.

Columns (2) and (4) add bibliometric controls measuring patent-pair similarity. We control for the number of shared technology (3-digit) classes, the number of shared (6-digit) subclasses, and the number of shared backwards citations to prior art. Most of the estimates of the effect of proximity on interference are robust to the inclusion of these controls and are precisely estimated. The effect of distance attenuates by about one-quarter when controlling for our measures of patent-pair similarity (panel A). Similarly, the effect of co-location within 161 km attenuates by about

Table 3: Robustness of co-location effect on interference

	3-digit controls			6-digit controls		
	μ [σ]	(1)	(2)	μ [σ]	(3)	(4)
A. Distance between pair of inventors						
Log distance	8.386	-0.008^c	-0.006^c	8.171	-0.041^c	-0.033^c
	[1.752]	(0.001)	(0.001)	[1.925]	(0.007)	(0.007)
No. shared classes	0.812		-0.012^c	1.141		0.008
	[0.662]		(0.003)	[0.844]		(0.022)
No. shared subclasses	0.130		0.166^c	0.788		0.286^c
	[0.502]		(0.015)	[1.078]		(0.033)
No. shared citations	0.011		0.099^a	0.056		0.130
	[0.631]		(0.057)	[1.676]		(0.096)
B. Co-location with same place, town, or city of residence						
1(Co-located in same place)	0.009	0.060^c	0.014	0.014	0.115	-0.042
	[0.092]	(0.020)	(0.023)	[0.118]	(0.112)	(0.126)
No. shared classes	0.812		-0.012^c	1.141		0.007
	[0.662]		(0.003)	[0.844]		(0.022)
No. shared subclasses	0.130		0.167^c	0.788		0.289^c
	[0.502]		(0.015)	[1.078]		(0.033)
No. shared citations	0.011		0.100^a	0.056		0.131
	[0.631]		(0.057)	[1.676]		(0.097)
C. Co-location with places of residence within 161km (100mi)						
1(Co-located within 161km)	0.057	0.043^c	0.030^c	0.077	0.206^c	0.156^c
	[0.232]	(0.007)	(0.007)	[0.266]	(0.048)	(0.048)
No. shared classes	0.812		-0.012^c	1.141		0.007
	[0.662]		(0.003)	[0.844]		(0.022)
No. shared subclasses	0.130		0.166^c	0.788		0.287^c
	[0.502]		(0.015)	[1.078]		(0.033)
No. shared citations	0.011		0.099^a	0.056		0.131
	[0.631]		(0.057)	[1.676]		(0.096)
Pairs	5,712,342	5,712,342	5,712,342	604,828	604,828	604,828
Pair-groups		831	831		821	821

This table shows estimates from a regression of an indicator for interference on a measure of proximity and controls. The sample includes interfering pairs of inventing teams and 3- or 6-digit controls, as indicated by the column group heading. The dependent variable is an indicator for interference × 100, with mean 0.04 percent in the sample with 3-digit controls and mean 0.4 percent in the sample with 6-digit controls. Sample means and standard deviations are reported in the columns labeled $\mu[\sigma]$. Panel A reports estimates where the measure of proximity is the minimum log distance between the pair of inventing teams. Panel B reports estimates where the measure of proximity is an indicator for whether or not the pair of inventing teams is co-located in the same place, town, or city of residence. Panel C reports estimates where the measure of proximity is an indicator for whether or not the pair of inventing teams is co-located with places of residences within 161km or 100mi. Standard errors, clustered on pair groups, are reported in parentheses. ^{a}—$p < 0.10$, ^{b}—$p < 0.05$, ^{c}—$p < 0.01$. The number of observations reported in the bottom panel applies to all regressions in that column.

one-quarter when including controls (panel C). These results suggest that, at least to the extent that we are able to introduce additional controls for factors contributing to the geography of invention, our results do not seem to be driven by imperfectly matched controls.

The effect of co-location in the same city, town or place (panel B) attenuates more when controlling for shared classes and citations but is estimated imprecisely. One explanation may be that measuring geographic proximity may be imprecisely measured using place of *residence*. One might want to use place of *work* or other places where inventors might be exposed to localized knowledge spillovers. However, we observe only place of residence. Relatedly, we also note that place of residence may refer to a very small spatial scale: a suburban town or village, or even an unincorporated community. At this spatial scale, other factors such as local amenities may contribute more to the sorting of inventors to places compared with localized knowledge spillovers. Finally, as already discussed, the co-location results are subject to bias related to spatial aggregation. For these reasons, we place more weight on the distance-based results in panel A or the co-location results at the spatial scale of a commuting zone in panel C.

Finally, we note that geography appears to contribute to interference over and above the number of shared citations. If all knowledge spillovers were recorded in citations, then controlling for citations should remove the effect of distance on interferences. Instead, our results suggest that proximity matters for localized spillovers of knowledge not recorded in citations, or tacit knowledge. We further explore this idea in the next section.

5 Comparison with citation-linked inventors

Are spillovers of *tacit* knowledge more localized than other forms of knowledge that are more easily codified? Arzaghi and Henderson's (2008) results suggest that the external benefits to advertising agencies in Manhattan attenuate quickly over space—they dissipate in as little as 750 meters. To the extent that interferences can capture spillovers of both tacit and codified knowledge, their localization could provide evidence that tacit knowledge spillovers require even closer proximity. To test this hypothesis, we compare the observed distribution of geographic proximity between interfering inventors to pairs of control patents and interfering patents linked by citation. Thus, our *control pairs* in this exercise are the *treated pairs* in Jaffe et al. (1993). As citation-linked patents have been used extensively following Jaffe et al. (1993) as evidence themselves of knowledge spillovers, we view these controls as a particularly strong counterfactual. For each interfering pair, we identify potential controls as patents *cited by* one of the interfering parties. Then, we form cited-citing control pairs by matching an interfering application to one of the these cited control patents.

This exercise also suggests that our main result is robust to matching controls on other factors. Here, matched control patents are selected based on citation links, rather than technology class and date of invention as in the main analysis presented in Section 3. Citation-linked patents seem even more likely to share similar production and demand factors, in addition to shared knowledge as measured by citations.

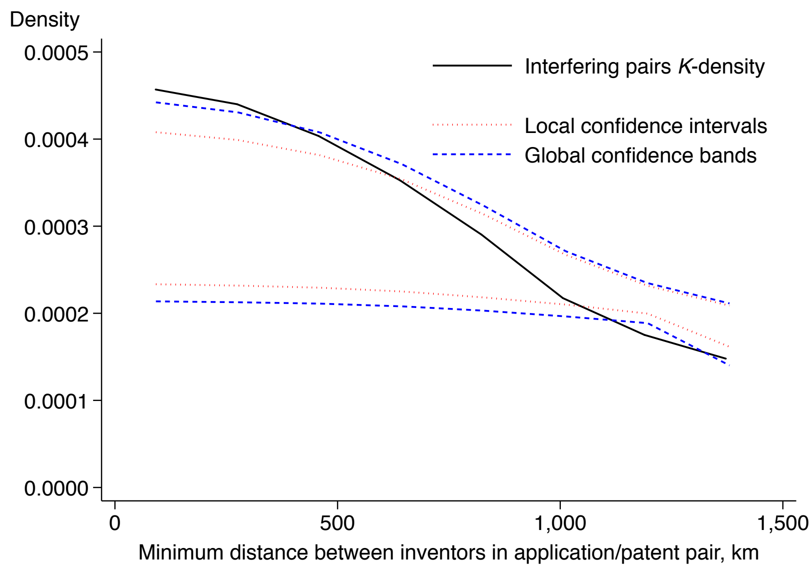


Figure 2: Distribution of geographic distances vs. citation-linked controls

This graph compares the estimated kernel densities of geographic distances between pairs of interfering inventors to the distribution for control pairs that include one interfering patent and one patent cited by the interfering patent. We use the minimum geodesic distance between the places of residence of inventors from opposite teams. The black line shows the estimated kernel density function for all interfering pairs for which we were able to find suitable controls. The dotted red lines show the upper and lower 5% local confidence intervals of the kernel density for non-interfering control pairs. The dotted blue lines show the upper and lower 5% global confidence bands of the kernel density for non-interfering control pairs. Control patents are restricted to backwards citations from one of the interfering patents. Interfering pairs are considered localized if the estimated kernel density of interfering inventors is above the upper global confidence band for at least one distance d up to the sample median.

Figure 2 shows the comparison between interfering inventor pairs and cited-citing control pairs. The black line again shows the observed density of pairwise distances between interfering inventors. (This is the same density reported in Figure 1, except for the change in sample to interferences with eligible *citation-linked* controls.) The dotted and dashed lines show local confidence intervals and global confidence bands, respectively, for the density of pairwise distances between cited-citing control pairs. Even compared against the geography of citation-linked patents, interfering patents are significantly more localized. A natural interpretation is that forms of input knowledge not captured easily by citations—perhaps tacit knowledge—contribute to the localization of patent interferences over that of patent citations. Thus, this result is consistent with the Jaffe et al. (1993) conjecture that citations are a *lower bound* on the strength of localized knowledge spillovers.

6 Localization and co-inventor ties

How is localized knowledge transmitted? Recent work suggests that network ties mediate the relationship between geographic proximity and localized knowledge spillovers. Social and professional ties may be especially important for the transmission of tacit knowledge. For example, Breschi and Lissoni (2009) use past co-inventorship as recorded on U.S. patents as a measure of social ties. They found that controlling for this measure greatly reduced estimates of the effect of geographic localization on citations. More recently, Head, Li, and Minondo (forthcoming) find that controlling for measures of the network ties among mathematicians halves the estimated effect of geographic distance on citations.

To shed light on the role of social ties in mediating the relationship between localization and interference, we follow Breschi and Lissoni (2009) and consider previous co-inventorship as a proxy for social ties. We use the Lai et al. (2013) database of inventors of U.S. patents to construct measures of ties. We define a network with each inventor represented as a node and connections or edges between any inventors that have been co-inventors on an issued patent. (Unique inventors are identified with name disambiguation, available in the Lai et al. (2013) database. Their algorithm uses not only name similarity but also inventor location, assignee, and technological class information.) The network distance between two inventors is the minimum path distance between them in the network—the number of edges along the shortest path from one inventor node to the other. This network distance is conditioned on co-inventor links up to 5 years *before* the earliest application date in interference. For a pair of patents A and B, we assign the shortest network distance between any inventor on patent A and any inventor on patent B.

An important limitation of this database is that it includes only information about co-inventor ties from issued U.S. patents. For interfering inventors without an issued patent—about one-quarter of inventors—we are unable to include them in this analysis, since they do not appear in the Lai et al. (2013) database. Further, social and professional ties outside of co-inventorship are not observed.

Interfering inventor pairs are more than twice as likely to be connected by previous co-inventor ties. Over two-fifths of interfering inventor pairs are connected by previous co-inventor ties, compared with about one-fifth of control inventor pairs.[31] This result confirms the findings of earlier studies suggesting that social and professional ties are important for facilitating knowledge flows.

However, we find only modest evidence that the localization of interfering inventors is accounted for by previous co-inventor ties. Table 4 shows the result of interference-rate regressions conditioned on co-inventor ties. Compared with Table 3, there are two more controls: (i) an indicator if two inventing teams *are not* connected by a previous co-inventor tie ($1(Network|observed) = 0$), and (ii) an indicator if two inventing teams *are* connected by a previous co-inventor tie ($1(Network|observed) = 1$). Unobserved network ties (for inventors outside the Lai et al. (2013) database) are the omitted

[31] See Appendix D.

category. Column (1) repeats estimates displayed in Table 3, column (3) for comparison. Columns (2)-(3) condition on previous co-inventor ties. Column (3) adds controls for patent-pair similarity as introduced in Table 3, column (4).

A few patterns stand out. One, there appears to be an interference premium for those connected by previous co-inventor ties compared with those that are not. The estimates in columns (2)-(3) of both panels suggest that pairs connected by previous co-inventor ties are about 0.06–0.07 percent more likely to be in interference compared with pairs not connected by previous co-inventor ties. This gap is robust to measures of geographic proximity and additional controls for patent pair similarity.

Two, pairs where network information is not observed are much more likely to be in interference, by about 0.5–0.6 percent. This pattern is accounted for by the fact that we do not observe network information for inventing parties who do not have an issued patent. Because control patents are selected based on having an issued patent, by construction, only the network information for interfering inventors is censored.

Three, the estimated effect of proximity does not change when we condition on observed co-inventor ties. In contrast to Breschi and Lissoni (2005), we find little evidence that the localization of interfering inventors is mediated via social ties, at least as proxied by previous co-inventor ties. These regression results suggest that this measure of social ties is independent of geographic localization. Of course, it is worth reiterating that an important limitation is that we do not observe other ties not captured by co-inventorship on U.S. patents. For only about two-thirds of the sample do we observe co-inventor ties, so part of the effect of log distance is identified by the one-third of the sample where we do not observe co-inventor ties.

7 Conclusions

We present new evidence of localized knowledge spillovers using a novel database of patent interferences—instances of simultaneous, identical invention by multiple, independent parties. By evidence of common, identical invention, interfering inventors share common knowledge inputs. Interfering inventor pairs show significant geographic localization compared with the counterfactual of inventor pairs sharing similar invention dates and technology classification. Thus, our results provide verification of the existence of localized knowledge spillovers and are distinct from the literature using patent citations. Interfering inventor pairs are even more localized compared with cited-citing inventor pairs, consistent with the Jaffe et al. (1993) conjecture that citations are a lower bound on the strength of localized knowledge spillovers.

Our results suggest that, in contrast to conventional wisdom about "the death of distance," geographic distance continues to matter, especially for flows of tacit, or difficult to codify, forms of knowledge. These are the types of knowledge flows where the lack of a "paper trail" has hampered the availability of evidence. Interferences therefore provide a unique and useful window

Table 4: Co-location and co-inventor ties

A. Distance between pair of inventors				
	μ [σ]	(1)	(2)	(3)
Log distance	8.171	-0.041^c	-0.040^c	-0.032^c
	[1.925]	(0.007)	(0.007)	(0.007)
No. shared classes	1.141			0.032
	[0.844]			(0.023)
No. shared subclasses	0.788			0.292^c
	[1.078]			(0.034)
No. shared citations	0.056			0.132
	[1.676]			(0.096)
1(Network|observed)=0	0.318		-0.539^c	-0.634^c
	[0.466]		(0.085)	(0.104)
1(Network|observed)=1	0.326		-0.471^c	-0.585^c
	[0.469]		(0.082)	(0.103)
R^2		0.065	0.066	0.070
Pairs	604,828	604,828	604,828	604,828
Pair-groups		821	821	821

B. Co-location with places of residence within 161km (100mi)				
	μ [σ]	(1)	(2)	(3)
1(Co-located within 161km)	0.077	0.206^c	0.195^c	0.146^c
	[0.266]	(0.048)	(0.048)	(0.049)
No. shared classes	1.141			0.031
	[0.844]			(0.023)
No. shared subclasses	0.788			0.293^c
	[1.078]			(0.034)
No. shared citations	0.056			0.132
	[1.676]			(0.096)
1(Network|observed)=0	0.318		-0.543^c	-0.637^c
	[0.466]		(0.085)	(0.104)
1(Network|observed)=1	0.326		-0.466^c	-0.582^c
	[0.469]		(0.082)	(0.104)
R^2		0.065	0.066	0.070
Pairs	604,828	604,828	604,828	604,828
Pair-groups		821	821	821

This table shows estimates from a regression of an indicator for interference on a measure of proximity and controls. The sample includes interfering pairs of inventing teams and 6-digit controls. The dependent variable is an indicator for interference $\times$ 100, with mean 0.4 percent. Sample means and standard deviations are reported in the column labeled $\mu[\sigma]$. Panel A reports estimates where the measure of proximity is the minimum log distance between the pair of of inventing teams. Panel B reports estimates where the measure of proximity is an indicator for whether or not the pair of inventing teams is co-located with places of residences within 161km or 100mi. Standard errors, clustered on pair groups, are reported in parentheses. ^{a}—$p < 0.10$, ^{b}—$p < 0.05$, ^{c}—$p < 0.01$.

into localized knowledge spillovers. In future work, it would be useful to leverage the potential of interferences to measure shared knowledge inputs to investigate other features of the invention process.

Works cited

Allen, Thomas J. (1984). *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization.* Cambridge, MA: MIT Press.

Arzaghi, Mohammad, and J. Vernon Henderson (2008). "Networking off Madison Avenue," *The Review of Economic Studies*, 75 (4): 1011–1038.

Audretsch, D.B., and M.P. Feldman (2004). "Knowledge Spillovers and the Geography of Innovation," *Handbook of Regional and Urban Economics*, vol. 4, 2713–2739.

Autor, D.H., and D. Dorn (2013). "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review* 103 (5): 1553–1597.

Baruffaldi, Stefano, and Julio Raffo (2017). "The Geography of Duplicated Inventions: Evidence from Patent Citations," *Regional Studies*, 51 (8): 1232–1245.

Bikard, Michaël (2012). "Simultaneous Discoveries as a Research Tool: Method and Promise," working paper.

Bleakley, H., and J. Lin (2012). "Thick-Market Effects and Churning in the Labor Market: Evidence from U.S. Cities," *Journal of Urban Economics*, 72 (2-3): 87–103.

Breschi, Stefano, and Francesco Lissoni (2005). "'Cross-Firm' Inventors and Social Networks: Localized Knowledge Spillovers Revisited," *Annals d'Économie et de Statistique* 79–80: 189–209.

Breschi, Stefano, and Francesco Lissoni (2009). "Mobility of Skilled Workers and Co-Invention Networks: An Anatomy of Localized Knowledge Flows," *Journal of Economic Geography* 9 (4): 439–468.

Calvert, Ian A. (1980). "An Overview of Interference Practice," *Journal of the Patent Office Society* 62: 290–307.

Calvert, Ian A., and Michael Sofocleous (1982). "Three Years of Interference Statistics," *Journal of the Patent Office Society* 64 (12): 699–707.

Calvert, Ian A., and Michael Sofocleous (1986). "Interference Statistics for Fiscal Years 1983 to 1985," *Journal of the Patent and Trademark Office Society* 68 (8): 385–393.

Calvert, Ian A., and Michael Sofocleous (1989). "Interference Statistics for Fiscal Years 1986 to 1988," *Journal of the Patent and Trademark Office Society* 71: 399-410.

Calvert, Ian A., and Michael Sofocleous (1992). "Interference Statistics for Fiscal Years 1989 to 1991," *Journal of the Patent and Trademark Office Society* 74: 822–826.

Calvert, Ian A., and Michael Sofocleous (1995). "Interference Statistics for Fiscal Years 1992 to 1994," *Journal of the Patent and Trademark Office Society* 77: 417–422.

Carlino, Gerald A., and William R. Kerr (2015). "Agglomeration and Innovation," in *Handbook of Regional and Urban Economics* vol. 5, Gilles Duranton, J. Vernon Henderson, William C. Strange (Eds.), Elsevier, 349–404.

Cohen, Linda R., and Jun Ishii (2006). "Competition, Innovation, and Racing for Priority at the U.S. Patent and Trademark Office," working paper.

De Simone, Daniel V., James B. Gambrell, and Charles F. Gareau (1963). "Characteristics of Interference Practice," *Journal of the Patent Office Society* 45: 503–591.

Duranton, Gilles, and Henry Overman (2005). "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies* 72: 1077–1106.

Duranton, Gilles, and Diego Puga (2004). "Micro-Foundations of Urban Agglomeration Economies," in J. Vernon Henderson and Jacques-François Thisse (Eds.), *Handbook of Regional and Urban Economics*. Amsterdam: Elsevier, vol. 4, chap. 48, 2063–2117.

Feldman, M. P. (2000). "Location and Innovation: The New Economic Geography of Innovation, Spillovers, and Agglomeration" in *The Oxford Handbook of Economic Geography*, G. Clark, M. Feldman and M. Gertler (Eds.). Oxford: Oxford University Press, 373–394.

Ganguli, Ina (2015). "Immigration and Ideas: What Did Russian Scientists 'Bring' to the United States?" *Journal of Labor Economics*, 33 (S1), U.S. High-Skilled Immigration in the Global Economy (Part 2): S257–S288.

Gladwell, Malcolm (2008). "In the Air: Who Says Big Ideas Are Rare?" *The New Yorker*, May 12, 2008.

Glaeser, Edward (1999). "Learning in Cities," *Journal of Urban Economics* 46 (2); 254–277.

Head, Keith, Yao Amber Li, and Asier Minondo (forthcoming). "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics," *Review of Economics and Statistics*.

Helmers, Christian and Henry Overman (2017). "My Precious! The Location and Diffusion of Scientific Research: Evidence from the Synchrotron Diamond Light Source," *Economic Journal* 127 (604): 2006–2040

Henderson, Rebecca, Adam Jaffe, and Manuel Trajtenberg (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment." *American Economic Review*, 95 (1): 461–464.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty (2000). "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review* 90 (2): 215–218.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108 (3): 577–598.

Kingston, William (2001). "Light on Simultaneous Invention from U.S. Patent Office 'Interference' Records," *World Patent Information* 26: 209–220.

Krugman, Paul (1991). *Geography and Trade*. Cambridge: MIT Press.

Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming (2013)."Disambiguation and Co-Authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)," `http://hdl.handle.net/1902.1/15705`. The Harvard Dataverse Network [Distributor] V5 [Version].

Lin, Jeffrey (2011). "Technological Adaptation, Cities, and New Work," *Review of Economics and Statistics*, 93 (2): 554–574.

Lin, Jeffrey (2014). "The Paper Trail of Knowledge Transfers," Federal Reserve Bank of Philadelphia *Business Review* Q2: 1–6.

Marshall, Alfred. *Principles of Economics*. London: Macmillan, 1920.

Merton, Robert K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations.* Chicago: University of Chicago Press.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura (2014). "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," *Review of Economics and Statistics* 96 (5): 967–985.

Ogburn, William F., and Dorothy Thomas (1922). "Are Inventions Inevitable? A Note on Social Evolution," *Political Science Quarterly* 37 (1): 83–98.

Pollack, Andrew (2017). "Board Rules on Patents for Process to Edit DNA." *The New York Times*, February 16, page B3.

Rosenbaum, Paul R. (2002). *Observational Studies*, 2nd ed. New York: Springer-Verlag.

Rosenthal, Stuart S., and William C. Strange (2001). "The Determinants of Agglomeration," *Journal of Urban Economics* 50 (2): 191–229.

Schmookler, Jacob (1966). *Invention and Economic Growth*, Harvard University Press.

Silverman, Bernard W. (1986). *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall.

Thompson, Peter (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, Peter, and Melanie Fox-Kean (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review* 95 (1): 450–460.

United States Patent and Trademark Office. "e-FOIA Reading Room," `http://e-foia.uspto.gov/Foia/PTABReadingRoom.jsp`, accessed November 3, 2014.

United States Patent and Trademark Office. "Efiling for Patent Trial and Appeal Board," `https://acts.uspto.gov/ifiling/PublicView.jsp`, accessed May 6, 2015.

United States Patent and Trademark Office. "Master Classification File," `http://patents.reedtech.com/classdata.php`, accessed January 16, 2014.

United States Patent and Trademark Office. "Public Patent Application Information Retrieval," `https://portal.uspto.gov/pair/PublicPair`, accessed June 11, 2013.

Weitzman, Martin L. (1998). "Recombinant Growth," *Quarterly Journal of Economics* 113 (2): 331–360.

Appendix

A Interference decisions

We downloaded 1,329 patent interference final decisions issued between 1998 and 2014 from the USPTO's "e-FOIA Reading Room." The decisions typically report the: (i) names of the interfering inventors; (ii) seniority status of each party; (iii) associated patent and application numbers; (iv) assignees; (v) judges' names; (vi) application claims in interference; (vii) decision on priority at the claim level (if there was one) or other disposition of the case; (viii) legal counsel; and (ix) hearing and decision dates. Sometimes, terse decisions omit some of these details. When available, we collect these details using additional documents found on the USPTO's "eFile" site or the Patent Application Information Retrieval (PAIR) service. The eFile site sometimes lists the notice declaring the interference, from which we can observe (x) inventors' location of residence. For cases with documents available on the eFile site, we also record (xi) notices of settlement agreements. Parties could also settle at any stage. Normally, details of these agreements were kept secret. These notices acknowledge the existence of a settlement agreement, as opposed to a decision on priority or some other outcome.[32] The PAIR service provides an alternate source of information on assignees, case disposition, the decision date, and inventors' locations.[33] Note that for inventors never (eventually) issued a patent, information on inventor location is available *only* on the notice of interference on the eFile site or the PAIR service.

B Claims and decisions

Table 5 displays statistics on claims of invention by parties according to their seniority status. (Recall that seniority status is typically determined by first-to-file and that the burden of proof in showing earlier conception and reduction to practice is on the junior party.) Panel A shows statistics for interference cases where we are able to observe both the claims made on all involved applications and the claims in interference. Panel B shows statistics for interference cases where we are able to observe only the claims in interference.

Interference counts correspond to most of application claims. The first row of Panel A shows that junior parties tend to make about 3 more claims on average compared with senior parties. (This difference in means is significant at the 5 percent level.) Of the 26 and 23 claims, respectively, made by junior and senior parties on average, 20 and 19 claims are declared in interference. (For the sample for which we only observe interfering claims, slightly fewer claims, 18 and 17, are declared in interference.) Thus, interferences tend to involve mostly identical competing claims of invention.

Senior parties tend to win two-thirds of the time. For the sample described in Panel A, junior parties lose 13 out of the 20 claims in interference, or 65 percent.

Partial decisions are rare. Out of the 688 junior parties, 465 (68 percent) lost *all* of their claims declared in interference. Out of the 687 senior parties, 221 (32 percent) lost all of their claims declared in interference. (The difference in these rates is significant at the 1 percent level.) These rates are similar to the larger sample described in Panel B (for which we observe only claims in interference). Overall, the patterns are also similar

[32] Unfortunately, settlement agreements are sealed. Thus, we can note their existence, but we cannot analyze their contents.

[33] In nearly every case where they overlap, information available on the PAIR record confirms information recorded from the decision.

to those seen in the subsamples of cases decided on priority or cases conceded. Thus, winners and losers tend to win or lose all of the claims in interference.

	All cases Jr.	*All cases* Sr.	*Priority decisions* Jr.	*Priority decisions* Sr.	*Conceded* Jr.	*Conceded* Sr.
A. Application claims observed						
Claims in application(s)	26.1	22.8^b	26.7	24.1	24.7	21.1^b
and/or patent(s)	(27.0)	(26.6)	(35.9)	(31.4)	(22.0)	(24.1)
Claims in	20.0	18.8	21.0	19.3	18.1	17.6
interference	(21.5)	(22.5)	(21.5)	(24.4)	(16.8)	(20.8)
Claims lost	13.3	6.6^c	17.9	3.9^c	12.8	5.0^c
in decision	(19.4)	(16.5)	(20.6)	(12.5)	(17.4)	(11.1)
Number of parties	688	687	121	116	358	360
Lost all appl./pat. claims	276	173	73	14	131	89
Lost all interfering claims	465	221	101	20	251	114
B. Only interfering claims observed						
Claims in	17.6	17.2	18.8	17.7	16.4	16.5
interference	(18.9)	(20.0)	(19.8)	(20.5)	(15.3)	(18.8)
Claims lost	12.7	6.5^c	15.9	4.1^c	12.0	6.0^c
in decision	(17.3)	(15.2)	(19.5)	(11.7)	(15.2)	(12.1)
Number of parties	1,236	1,102	257	214	627	563
Lost all interfering claims	910	405	212	47	469	221

Table 5: Claims for senior and junior parties

This table reports means and standard deviations (in parentheses) for senior and junior parties. Seniority is determined before an interference proceeding begins, according to the earliest benefit date on file. (Typically, the benefit date is the earliest date of application to the USPTO or a foreign patent authority.) Number of claims is the sum of claims across all applications filed by each independent interfering party. H_0: Difference in means by seniority is zero. ^{a}—$p < 0.10$; ^{b}—$p < 0.05$; ^{c}—$p < 0.01$.

C Shared codified knowledge inputs

Interfering inventors tend to share common knowledge inputs. In this section, we provide evidence of shared knowledge inputs among interfering and control pairs by analyzing two commonly-used bibliometric measures of patents and applications – shared technology classifications and shared backward citations. These results connect interferences to more familiar measures of shared knowledge. They also motivate our analysis conditioning on additional controls.

Interfering patents and applications have a similar number of total technology classifications and subclassifications compared with control patents. Recall that patents may be classified with multiple technology

	Interfering	3-digit controls μ	C.I.	6-digit controls μ	C.I.
A. Means for applications/patents					
Backwards citations	10.3	11.6	(10.7, 12.5)	11.5	(10.6, 12.4)
USPC classes	1.92	2.28	(2.22, 2.35)	2.23	(2.17, 2.29)
USPC subclasses	4.99	5.97	(5.69, 6.27)	7.01	(6.64, 7.42)
B. Means for pairs of applications/patents					
Backwards citations shared	3.40	0.03	(0.00, 0.07)	0.37	(0.20, 0.76)
USPC classes shared	1.28	0.86	(0.83, 0.89)	1.04	(1.01, 1.07)
USPC subclasses shared	1.58	0.11	(0.09, 0.14)	0.71	(0.66, 0.76)

Table 6: Interfering inventors share codified knowledge inputs

This table compares means of interfering patent pairs to simulated means, 5th- and 95th-percentile estimates for control pairs. A control pair includes 1 interfering application and 1 issued control patent that share a technology class and application date. Simulated CIs based on 1,000 random draws from eligible control pairs. The sample is cases with decisions on priority and concessions.

classifications and sub-classifications when their subject matter overlaps with multiple technology areas. Table 6 displays summary statistics for interfering applications or patents and associated controls. Panel A compares the total number of citations and technology classifications between interfering and control patents and applications. The first column shows means for interfering patents and applications. The subsequent columns compare this statistic to corresponding statistics for control patents. We report means and lower 5% and upper 5% confidence levels from our Monte Carlo simulations for eligible 3- and 6-digit control patents.

Patents must cite other patents as prior art, so that the number and type of "backwards citations" can then be derived from the citations made by a patent. Interfering patents and applications have a similar number of total backwards citations, or citations to previously issued patents as "prior art," compared with control patents. On average, interfering patents and applications contain 10.3 backwards citations. Control patents have about one more citation, at 11.5.

Interfering patents and applications tend to have slightly fewer U.S. patent classifications and subclassifications compared with control patents matched on 3-digit technology classes and 6-digit subclasses. These differences are statistically significant, but small in magnitude, 1.92 patent classifications for interfering versus 2.28 and 2.23 for control patents. For sub-classifications, the means are 4.99 for interfering patents and applications and 5.97 and 7.01 for the controls. These small differences are entirely accounted for by the rate of interfering applications—about a quarter—that are never passed for issue. For these applications we observe only one primary class and subclass in the PAIR database.[34]

Overall, interfering patents and control patents are similar in terms of the total number of backwards citations to prior art and the total number of technology classes and subclasses. These results support the validity of the matched control strategy, since these factors were not used in selecting matched controls.

However, interfering pairs differ in terms of the number of *shared* backwards citations and technology

[34] If we compare only primary classes and subclasses, both interfering and control pairs have on average 1 primary class and subclass, by construction. The actual number of primary classes and subclasses for interfering pairs is slightly more than 1 owing to a small number of interfering parties with claims spread across multiple patents or applications.

classes and subclasses. This is an indicator that interfering pairs overlap more in terms of shared knowledge inputs, as measured by these common bibliographic measures of patents. Panel B reports statistics for the number of backwards citations and technology classifications shared within interfering pairs and control pairs. Despite the overall similarity in the total number of backwards citations and technology classes and subclasses, interfering pairs share substantially more backwards citations and technology classes and subclasses. Interfering pairs share 3.4 backwards citations, compared with 3-digit control pairs that share 0.03 backwards citations and 6-digit control pairs that share 0.37 backwards citations. Interfering pairs also share more U.S. patent classifications. Interfering pairs tend to share 1.28 classes and 1.58 subclasses. This is more than the number of classes and subclasses shared by both 3- and 6-digit control pairs. As noted earlier, for about a quarter of interfering applications we observe only one primary class and subclass. If we focus on primary classes and subclasses, 62 percent of interfering classes share a primary class and 25 percent share a primary subclass, compared with 48 percent and 7 percent of 6-digit control pairs and 35 and 1 percent of 3-digit control pairs, respectively. Thus, the result that interfering pairs share more classes and subclasses is robust to conditioning on primary classes and subclasses, for which we have complete data. Overall, these results suggest that interfering inventing teams may share more *codified* knowledge inputs compared with control pairs.[35]

D Co-inventor ties

Interfering inventor pairs are more than twice as likely to be connected by previous co-inventor ties. Over two-fifths of interfering inventor pairs are connected by previous co-inventor ties, compared with about one-fourth of control inventor pairs (see Figure 3). This result confirms the findings of earlier studies suggesting that social and professional ties are important for facilitating knowledge flows.

[35]In Sections 3 and 5, we also check that our results are robust to conditioning on these bibliometric measures of similarity.

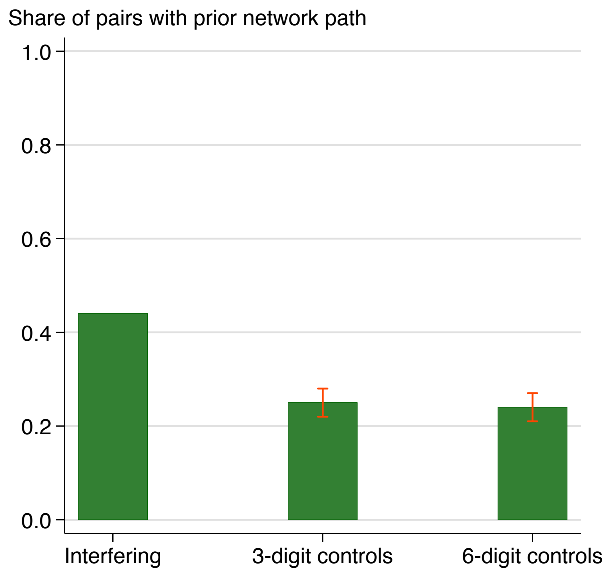


Figure 3: Interfering inventors are more likely to be connected by previous co-inventor ties

Priority decisions only. Simulated 90 percent confidence intervals shown.