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

Ina Ganguli† Jeffrey Lin‡ Nicholas Reynolds§

May 28, 2015

Abstract

We present new evidence of localized knowledge spillovers using a novel database of patent interferences—instances of simultaneous, identical invention by multiple, independent parties.

Keywords:
JEL classification:

*Super preliminary, please do not cite. We thank Jerry Carlino and Bob Hunt for comments and suggestions, and Aaron Rosenbaum for excellent research assistance. The views expressed here are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

†Assistant Professor, Stockholm Institute of Transition Economics, Stockholm School of Economics.

‡Economic Advisor and Economist, Federal Reserve Bank of Philadelphia. E-mail: jeff dot lin at phil dot frb dot org.

§Graduate student, Brown University.

1 Introduction

Why do firms and inventors tend to locate in dense, costly areas? One intriguing hypothesis is that such geographic concentration lets them benefit from local knowledge spillovers. As Lucas (1988) noted, "what can people be paying Manhattan or downtown Chicago rents for, if not for being near other people?" Localized knowledge spillovers are central to theories of economic growth and the existence of cities. But despite their theoretical and anecdotal significance, evidence of their existence remains mixed. In addition, a key counterargument to the importance of knowledge spillovers is that firms might prefer to keep their work secret from competitors. Are cities really places where knowledge is "in the air"?

In this paper, we propose a new test of localized knowledge spillovers. We create a novel database of patent interferences—instances of simultaneous invention by multiple, independent parties—to provide new evidence on the role of knowledge spillovers in invention. By examining patent interferences, we highlight the role of common knowledge inputs in invention, versus the sharing of inventive output. Interferences especially emphasize shared tacit, or difficult to codify, knowledge inputs.

Until 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. When the U.S. patent office received applications from multiple, independent parties with identical claims at roughly the same time, it was obliged to investigate the competing claims to determine which party was entitled to patent protection. This investigation, known as a patent interference proceeding, determined who had first conceived of the invention and reduced it to practice. Typically, parties submitted dated laboratory notebooks, testimony by associates, and media reports as evidence of first invention.

Importantly, by recording instances of common invention, patent interferences create a valuable and unique record of common inputs. In particular, inventors in interference are likely to have command of similar knowledge inputs. In the spirit of Weitzman (1998), the view that new ideas result from combinations of existing ideas suggests that interfering inventors are likely to share access to the same existing knowledge. For example, interfering inventors may have similar background in chemistry, or they may have similar expectations of market conditions. Patent interferences are therefore especially valuable for capturing spillovers of tacit knowledge inputs that are difficult to measure in studies that rely on citations. Further, interference cases often reveal details about the inventive process that allow us to rule out alternative explanations—outside of shared knowledge inputs—for multiple invention. Thus, a patent interference suggests a likely knowledge spillover has occurred in the inventive process.

We construct a database of interference cases from the early 1980s to 2011. Many details, including the identities of participants and winners, are hand-collected from judges' final opinions. We merge these data with other sources containing additional information about patents, applications, and inventors. We use these data to study the geography of knowledge spillovers. Following the matched-control strategy of Jaffe, Trajtenberg, and Henderson (1993), we find that interfering inventors tend to be geographically closer to each other versus similar inventors working in the same year and technology field. Thus, inventors working independently but using common knowledge inputs are substantially more geographically concentrated than other similar inventors. Our results provide new evidence of localized spillovers of tacit knowledge.

2 Literature review

For Marshall (1920), knowledge in dense cities was difficult to keep secret: "The mysteries of the trade become no mysteries; but are as it were in the air." Today, localized knowledge spillovers are central to theories of cities (e.g., Davis and Dingel, 2013). Recent evidence seems to support this hypothesis. Carlino, Chatterjee, and Hunt (2007) found that patent intensity, or patents per capita, were positively correlated with urban density. Research and development labs are also geographically concentrated (Carlino, Carr, Hunt and Smith, 2012). A key challenge to interpreting this evidence is that many other factors might also encourage inventors and firms to locate near one another. For example, firms might benefit from better matching with specialized workers, or they might better exploit opportunities for sharing other production inputs. Or skilled inventors may be attracted by superior amenities offered by big cities.

Jaffe, Trajtenberg, and Henderson (1993) proposed a control-matching strategy to deal with this identification problem. They argued that knowledge flows could be tracked via patent citations—when on inventor cited important prior art as a knowledge input. Then, they compared the geographic distance between a "citing" and "cited" inventors to the geographic distance between the cited inventor and a matched "control" patent that was similar to the citing patent in terms of technological classification and date of application, except for the citation link. Thus, the control patent-cited patent link represents the expected proximity of inventors working in the same time period and research field, but not conditioned on a "knowledge spillover" (i.e., a citation). If the inventors of the citation-linked patent pair are observed in closer proximity than this benchmark, then this is strong evidence that a localized knowledge spillover has occurred, especially since we have accounted for the underlying geographic distribution of research activity and hence why inventors might be located together. Jaffe et al. found evidence along these lines. However, subsequent work, including Thompson (2006) and Thompson and Fox-Kean (2005), found that these results were fragile—choosing to exclude patent examiner-added citations, for example, or selecting control patents in a different way, eliminated the Jaffe et al. result. Recently, Murata et al. (forthcoming) propose a distance-based version of the Jaffe et al. test, based on methods in Duranton and Overman (2005).

They find evidence in support of localized patent citations.

Our study departs from these papers by dispensing with patent citations entirely. Instead, we focus on multiple identical inventions as indicators that inventors have shared common knowledge inputs. By examining patent interferences, we highlight the role of common knowledge inputs in invention, versus the sharing of inventive output. We also emphasize knowledge inputs other than those that can be codified and cited—i.e., sharing of "tacit" knowledge inputs. Such tacit knowledge may be even more sensitive to proximity, as noted by Jaffe et al. (1993). Shared knowledge inputs might also include information that is not patented—new scientific insights, for example, or information about potential demand.

Finally, multiple invention has been long studied in the history of science. Merton (1973) and Ogburn and Thomas (1922) examine multiple identical inventions and discoveries and contrast explanations based on the "heroic" theory of invention (i.e., the lone inventor) versus a "social" theory of invention (i.e., such inventions were inevitable). Rather than examine competing theories of multiple invention, in this paper we exploit the fact of identical inventions to identify instances of shared knowledge inputs.

3 Patent interferences

3.1 Background

Patent interferences were a unique feature of U.S. patent law. Through 2013, the U.S. had a "first to invent" rule for assigning priority of invention, versus a "first to file" rule more common in the rest of the world. When the U.S. patent office received patent applications from multiple, independent parties with one or more identical claims at roughly the same time, it was obliged to investigate the competing claims to determine which party was entitled to patent protection. This investigation, known as a patent interference proceeding, determined who had first conceived of the invention and reduced it to practice. Typically, parties submitted dated laboratory notebooks, testimony by associates, and media reports as evidence of first invention.

In this section, we summarize several key institutional features of patent interferences. 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).

Interferences were declared by a patent examiner when (i) at least two simultaneous U.S. patent applications or (ii) one U.S. patent application and a recently-issued patent contained identical claims of invention. (Patent examiners would discover interferences in their normal searches for prior art.) 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 U.S. P.T.O. 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.

In the interference proceeding, the examiner defined "counts" corresponding to the interference. Each count was characterized by a distinct invention at issue; each application might claim several distinct inventions so an interference might involve multiple counts. The case was then sent to a rotating three-judge panel from the Board of Interferences. Inventors were assigned a "benefit date"—typically, the date of application, either at the U.S. patent office or a foreign patent office. The inventor with the earlier benefit date was the senior party. The burden of proof—i.e., demonstration of an earlier conception and reduction to practice—was on the junior party. Interfering inventors typically submitted lab notebooks, eyewitness testimony, and other forms of independent corroboration to prove they were the first to invent. Finally, the judges would issue a decision on priority or report some other outcome.

Table 1 summarizes common interference dispositions. Parties could settle at any stage. Normally, details of these agreements were kept secret. If the case was not settled before the case went to the Board of Interferences, then chief judge would report the case disposition. (Note that settlements reached after the hearing stage would be reported by the Board.) One party might also concede the case, without a settlement taking place—for example, if they realized their case was weak.

A decision on priority meant a judgement that one party had first conceived of the invention and reduced it to practice.

Other potential outcomes involved specific rules of interference, and they are useful in distinguishing instances where inventors shared knowledge inputs versus other reasons for interference. For example, according to interference rules, an interfering inventor would lose priority if it they had not promptly filed for a patent application following conception and reduction to practice. 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. Decisions against interfering inventors who chose to keep their invention secret for some time suggest that the incentive to delay disclosing an invention promptly (by filing a patent application) might be mitigated by this rule. (In addition, interference cases appear to vary in terms of whether inventors are aware of each other's efforts.)

Sometimes, the Board of Interferences might decide that the examiner that declared the interference was mistaken, and that there was no interference in fact. This reported result is important because it allows us to isolate true multiples, and not just near misses.

Disposition	Example (X vs Y)	Patent to
Priority	Y conceived and reduced to practice first	Y
Settlement	X concedes; Settlement terms confidential	Y
Abandonment	X concedes; No settlement	Y
Concealment	X kept invention secret or concealed best mode	Y
Derivation	X stole invention from Y; i.e., prior conception by Y and communication of conception to X	Y
No interference in fact	Claims are actually distinct	X and Y
Common ownership	X and Y work for same conglomerate	X or Y
Anticipation	Inventors' claims were disclosed in an existing patent	No one
Unpatentable	Claims are not patentable (e.g., obvious)	No one
Inventorship	Special case: X+Y vs X where Y was co-inventor	X+Y

Table 1: Common interference dispositions

The U.S. P.T.O. worked hard to make sure that interfering inventors did not share other factors. Claims from parties working for the same firm (e.g., different branches of a large corporation) were dismissed. If one party could be shown that information about the invention had been disclosed by the competing party and communicated to them (e.g., stolen), then that was grounds for an unfavorable judgment.

Finally, in some less common outcomes, the Board ruled that the interference count was anticipated or otherwise unpatentable. As with no-interference-in-fact, these judgments could be interpreted as mistakes by the original patent examiner.

3.2 Interferences indicate common knowledge inputs

Importantly, by recording instances of common invention, patent interferences create a valuable and unique record of common inputs. In particular, inventors in interference are likely to have command of similar knowledge inputs. In the spirit of Weitzman (1998), the view that new ideas result from combinations of existing ideas suggests that interfering inventors are likely to share access to the same existing knowledge. For example, interfering inventors may have similar background in chemistry, or they may have similar expectations of market conditions. Patent interferences are therefore especially valuable for capturing spillovers of tacit knowledge inputs that are difficult to measure in studies that rely on citations.

Further, interference cases often reveal details about the inventive process that allow us to rule out alternative explanations—outside of shared knowledge inputs—for multiple invention. Thus, a patent interference suggests a likely knowledge spillover has occurred in the inventive process.

First, interferences were declared by a patent examiner who specializes in a particular technological area. Thus, interfering claims are likely to be detected. In some cases, the examiner was alerted to a possible interference by one of the applicants, but an interference is distinct from patent infringement, in which the holder of an existing patent sues an infringing party. In contrast to infringements, private parties could not sue for an interference.

Second, interferences must involve parties with simultaneous pending applications for patents.[1] This is important for several reasons. One, as noted

[1] 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).

in the previous paragraph, this feature makes interferences distinct from patent infringements, which typically involve leaders and followers. Two, the simultaneity requirement reduces the likelihood of copying or stealing—that is, that one inventor's claims are directly sourced from the competing party. Instead, it is more likely that both parties are drawing knowledge inputs from shared information. In addition, as mentioned earlier, evidence of stealing or espionage is grounds for adverse judgement in the interference decision.

Third, interferences are likely to involve valuable patents, and thus actual inventions. We show later that this appears to be true, at least in terms of forward citations.

Fourth, interferences seem unlikely to result from other factors in common or within-firm spillovers, since interferences between parties with common ownership were not allowed.

Fifth, as mentioned earlier, no-interferences-in-fact help us to distinguish between identical inventions and near misses.

Sixth, some interferences represent cases where patent applications completely overlap. Thus, contra Schmookler (1966), interferences can identify identical inventions, versus near misses. The P.T.O. tracks of the number of application claims corresponding to each interference count. Therefore, it is possible to separate interferences where all application claims are in interference, versus other cases where only some application claims are in interference.

3.3 Case study

An important recent interference (number 102,416) involved competing claims for the method of producing the Hepatitis-B vaccine from yeast. This case is somewhat unusual in that the patent was especially valuable and the case dragged through the court system for many appeals and years. However, many features of the case highlight important general features of interferences as indicators for shared knowledge inputs.

The case involved two competing teams. William Rutter, Pablo Valenquela, Benjamin Hall, and Gustav Ammerer were scientists at the University of California and University of Washington, whose work was in part funded by Merck. Ronald Hitzeman, Arthur Levinson, and Daniel Yansura were scientists at Genentech. Rutter et al. filed their application on August 4, 1981, claiming simultaneous conception and reduction to practice on June 30, 1981. Hitzeman et al. filed their application on August 31, 1981, claiming conception and February 3, 1981 (five months before Rutter) and reduction to practice on July 20 (three weeks after Rutter). After a lengthy set of motions and appeals, the case was finally decided by the U.S. Court of Appeals for the Federal Circuit (2001) in favor of Rutter, on the basis that Hitzeman's earlier claimed conception date was invalid. The court ruled that Hitzeman's actual conception date was July 20. Since Rutter's team now had an earlier conception and reduction to practice, they were awarded the patent.

Several features of this case are of interest. According to Kleid's (2002) oral history of the case, and the decision by the appeals court (2001), both teams had knowledge inputs in common. (The following discussion is drawn from these two sources.) It was already known that the Hepatitis-B antigen could be purified from the blood of certain infected humans. This so-called "Australia antigen" could then be used as a vaccine; this advance was awarded the 1976 Nobel Prize in Medicine, but the vaccine produced in this way remained costly to manufacture. Many teams, including the two in interference, speculated that a Hepatitis-B antigen might be produced using bacteria or yeast. However, previous attempts by other scientists to use E. coli to produce the antigen had failed.

In the late 1970s and in 1980, Hitzeman had collaborated with scientist John Carbon to successfully produce the "interferon" protein in yeast, providing a model for producing the Hepatitis-B antigen in yeast. After this work, he joined Genentech and collaborated with Hall and Ammerer on other projects involving interferon. Shortly thereafter, both teams decided to try using yeast to manufacture the antigen.

Thus, members of both teams had similar knowledge inputs: the market need for a low-cost Hepatitis-B vaccine; the costly "Australia antigen" method; the failed attempts using E. coli; and the successful use of yeast to produce other proteins.

4 Data and methodology

4.1 Interferences, patents, and applications

We construct a novel database of patent interference cases from the early 1980s to 2013. Our database starts with information hand-collected from interference decisions of the U.S. Board of Patent Appeals and Interferences. For each interference case, these decisions identify the inventors involved, their applications or patents, a narrative description of the case, and the disposition or judgment of the Board. We merge these data with other existing patent databases to match interference cases to patent citation links, technological classifications, the geographic location of inventors, and other characteristics. Notably, we also obtain the characteristics of non-issued patent applications (i.e., applications by losing parties in interference) from the Patent Application Information Retrieval (PAIR) service.

	Freq	**%**
Settlement or abandonment	784	59.7
Priority	270	20.5
Unpatentable	97	7.4
Common ownership	63	4.8
No-interference-in-fact	47	3.6
Other	53	4.0
Total	1,314	100.0

Table 2: Distribution of interference case dispositions

We start with a database of interference decisions, available on the USPTO Board of Patent Appeals and Interference's public, online "e-FOIA Reading Room." This database includes a table of 1,314 interference cases with final decisions issued between 1998 and 2014.[2] For each case, we review the judges' decisions and record information about the interference, including the parties involved and the ultimate disposition of the case. We select only interference cases where the Board's decisions report a settlement or judgment on priority (see Table 2). We expect that shared knowledge inputs are more likely in these cases. Sometimes, particularly terse final decisions omit certain key details, such as patent or application numbers associated with the interference. When available, we collect these details using additional documents found on the USPTO's "E-filing" site. From these sources, we record patent and application numbers, inventor names, companies to which the patents were assigned, the benefit dates granted for determination of priority and the resulting seniority status in the case, the type of decision reached, and the effect of the decision on each party's patent rights.

Next, we match these data on interference cases to patent and application data. Many interference cases occur between a party with an issued patent and another with only a pending patent application. If the interference is decided in favor of the party that already has an issued patent, then the losing party's patent application usually never becomes an issued patent. Since most patent databases used by economists only include data for issued patents, we must obtain information about denied applications from an alternative source. Thus, we use the PAIR data hosted by Google Patents, which contains application-level data for applications filed since 2001. We extract the name and location of the first-named inventor, the identities of the patent examiner and attorney, and the filing date for each application.

[2]There are a few decisions related to interferences declared well before 1998, as far back as the early 1980s, including the Hepatitis-B case noted earlier. 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.)

We also take advantage of the associated transaction history file to record all interference cases associated with the patent application, decisions dates, and whether the ruling was favorable or unfavorable.

For issued patents, we also match our interference case database to the inventor disambiguation dataset of Lai et. al (2013). For all patents issued by USPTO between 1975 and 2010, this includes the names and locations of all inventors, the patent application date, and assignee. This dataset also includes a unique inventor identifier that is consistent across all patents, as the result of name disambiguation algorithm. Importantly, this allows for the examination of the entire inventing careers of particular inventors and for the construction of a social network of inventors that allows for the measurement of network distance.

Finally, we obtain technology classification information directly from the USPTO's Master Classification File. This information is available for all issued patents, but only for patent applications starting on March 15th 2001. For earlier patent applications we instead obtain classification information from the PAIR bulk downloads file, which only includes the primary classification.

4.2 Geographic and social proximity

Many interfering inventors appear to be in close proximity. We construct the geographic distance between the inventors of the two patents in each pair. For issued patents we obtain the latitude and longitude of each inventor from the dataset of Lai et al. (2013). For patent applications we are only able to obtain the place and state of the first-named inventor from the PAIR downloads, and we match this to the Census Gazeteer file of place names to obtain latitude and longitude. We then calculate the crow-flies distance for each possible pairing of inventors. For each patent pair, we have a number of pairwise combinations of inventors, each potentially being a different distance apart, and we record the median, mean and minimum of these distances.

(Also, following the convention of earlier literature we record the distance between the first-named inventors of each patent.) Among interferences that have valid 6-digit classifications, 8.0% have interfering inventors that are no more than 100 kilometers apart.

Next, following Breschi and Lissoni (2005), we construct a measure of the "social network distance" between the inventors of the two patents in each pair. We define a social network with each inventor represented as a node and connections or edges between any inventors that have been co-inventors on an issued patent . We use the dataset of Lai et al. (2013), which has data on the inventors of all patents issued by the USPTO between 1975 and 2010. Importantly, the creators of this dataset use a name-disambiguation algorithm to identify unique inventors (using not only name similarity but also inventor location, assignee, and technological class information). This helps to prevent the creation of spurious inventor connections due to different inventors with similar names, or the lack of connections caused by inconsistent naming conventions by inventors. We construct a social network of inventors using the above definition and dataset. We define the social distance between two inventors as the minimum path distance between them in the network—the number of edges along the shortest path from one inventor node to the other. Following the convention described above for geographic distance, we first record the network distance between inventor pairs, and then for each patent pair we record the median, mean and minimum social distance between inventors, as well as the social distance between first-named inventors.

4.3 Simulation of counterfactual distances

Can localized knowledge spillovers be identified by the geographic concentration of inventive activity? Inventors might choose locations with other important factors (capital, skilled workers) or greater demand. Thus, it is unclear whether the co-location of inventors indicates that knowledge flows are empirically relevant. Jaffe et al. (1993) suggested that by comparing the distribution of citation-linked inventors to appropriately matched control patents, they could separately identify these controls.

We test for tacit knowledge flows via geographic and social proximity of interfering inventors, following the control-matching strategy of Jaffe et al. (1993), Breschi and Lissoni (2005), and Murata et al. (forthcoming). We compare patterns of interfering inventions to counterfactual patents—patents that share identical technological classifications and similar application dates. By using counterfactual patent distributions, we hope to control for all factors except common knowledge inputs. Thus, significant deviations from the counterfactual suggest that knowledge flows are empirically important.

We can easily estimate distributions and calculate means for our dataset of interfering pairs. Following Duranton and Overman (2005) and Murata et al. (forthcoming), we run simulations to allow us to compare these statistics in a formal way to a set of counterfactual pairs. We run 1000 simulations. In each simulation, we first randomly select a control pair for each interfering pair from the set of permissible control pairs defined below. Using this simulated draw of controls we estimate kernel density functions for the distribution of pairwise geographic and social proximity. Additionally, we calculate the means of a number of other statistics. We then draw a new set of control pairs and re-estimate k-densities and means for each new simulation, repeating a total of 1000 times.

After running 1000 simulations, we separately 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 95% confidence band. For the other statistics we rank the estimated means and select the 50th ranked mean as the 5% lower confidence estimate and the 950th as the 95% upper estimate.

Our constructed local confidence bands only allow us to make local statements about the relationship between the interfering and counterfactual distributions (i.e., if the interfering distribution is above the upper counterfactual band at 100 km we could say that interference pairs are localized at 100 km). By construction there is a 5% probability for each particular distance that a random draw of counterfactual pairs shows localization and thus across all distances a much higher likelihood of displaying localization at some distance. This effect is attenuated slightly by autocorrelation between the densities at different distances. However, clearly to make global statements about localization we need stricter criteria than the local confidence bands. For this task we construct global confidence bands, following the approach of Duranton and Overman (2005) and Murata et al. (2013). First we define d_a as the median distance of all counterfactual pairs. Then, we return to our simulated draws and search for an upper and a lower local confidence band such that, when we consider them across all distances between 0 and d_a, only 5% of our simulated K-densities hit them. In general, we end up selecting approximately the 99% local confidence band as the 95% upper global confidence band and the 1% local confidence band as the 5% lower global confidence band. Thus we set a higher bar for evidence of global localization than that for local localization.

4.4 Control patents

We construct control pairs in the following way. First, we reshape our database of interferences into pairs of interfering inventors. Since some interference cases involve more than two parties, our database of 1,314 interference cases involves 1,458 interfering pairs of inventors. For each pair, we record all 3- and 6-digit technology classes shared between the two parties. We also record the application date of each party.

Next, we select a set of control patents that are similar to the interfering patents, but are not involved in the interference case. Using the Lai et al. (2013) database and the USPTO master classification file, we search for controls in the entire universe of patents issued between 1972 and 2013. For each interfering patent or application, we select a set of controls based on two criteria. First, we require a control patent to have 3-digit technology classification that was shared by the interfering pair. Second, all control patents must have an application date that falls between the application dates of the two interfering parties. If no eligible controls are found, we then expand the application date "window" incrementally by ten days before the earlier interfering application and 10 days after the later interfering application, until an eligible control patent is found. Additionally, following Thompson and Fox-Kean (2005), we select a second, finer set of controls in which we restrict controls to share a 6-digit technology classification with the interfering patents.

Next, we match each of these control patents with each of the two interfering patents, creating two control pairs. Each control pair thus includes one interfering patent and one control patent that is technologically similar and has a similar application date. In the end, for each pair of interfering patents, we end up with a pool of control pairs that are constructed in a particular way to be similar to the interfering pair. Specifically, by construction, the application date difference between a control pair will be no more than 180 days greater than that of its associated interference pair, and the control pair will share a technological classification that is also shared by the interfering pairs.

See the appendix for more details on picking controls.

4.5 Results

4.5.1 Interfering vs. control pairs

Table 3 shows summary statistics of interfering pairs, as well as 5th and 95th percentile estimates for those means for pairs of interfering and control patents. Simulation (A) displays means for controls that share a 6-digit technology class with both interfering patents, as in Thompson and Fox-Kean (2005). Simulation (B) displays means for controls that are linked to an interfering patent by a backward citation. Thus, the counterfactual being considered is a knowledge spillover as evidenced by a citation link, a la Jaffe et al. (1993).

In simulation A, application dates for interfering pairs are similar to control pairs by construction. In simulation B, control pair application dates predate interfering pair dates because we condition on a backward citation to find a control patent.

Against both counterfactuals, interfering patents are more closely linked by standard patent measures. Interfering patents (i) share more backwards citations, (ii) share more 3-digit technological classifications, and (iii) share more 6-digit classifications. One way of interpreting these results is that interfering inventions are "closer" in technological space than either standard control-matched patents or even citation-linked control patents. In other words, these results are consistent with interferences identifying identical versus near-miss inventions, and sharing both codified and tacit knowledge inputs.

4.5.2 Geographic proximity

First, we test the hypothesis that interferences are more geographically concentrated versus similar inventive activity not linked by common invention. Figure 1 compares the distribution of geographic distances between pairs of interfering inventors to the distribution for similar, non-interfering pairs. 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 local confidence bands of the kernel-density for non-interfering, control pairs. The dotted blue lines show the global confidence bands of the kernel-density for non-interfering, control pairs. For each interfering and control pair we used the minimum pairwise distance among co-inventors. In Panel

Table 3: Interfering patents versus control patents

	Simulation (A)			Simulation (B)		
	Interfering	6-digit controls		Interfering	JTH controls	
Application date†	10/6/96	9/22/96	10/30/96	9/21/96	12/25/92	4/20/93
Δ Application date†	716.17	340.37	379.03	740.77	2450.74	2671.30
Total backwards citations	18.24	14.02	16.47	18.99	16.35	17.41
Citations shared	3.90	0.35	1.20	2.83	0.68	0.96
Number of 3-digit classes	2.02	2.07	2.16	1.97	2.04	2.11
Number of 6-digit classes	5.36	5.34	5.73	5.33	5.30	5.59
Shared 3-digit classes	1.53	1.30	1.37	1.38	1.12	1.19
Shared 6-digit classes	2.44	1.37	1.48	1.94	0.94	1.06
Number of pairs	504 pairs			575 pairs		

Notes: This table compares means of interfering patent pairs to simulated 5th- and 95th-percentile estimates of those means for interfering-control patent pairs. Simulation (A) displays means for controls that share a 6-digit technology class with both interfering patents. Simulation (B) displays controls that are linked by (backward) citation to one of the interfering patents. †–Application dates do not reflect "benefit" dates accorded to applicants on the basis of continuations or foreign patent applications. Benefit dates are typically earlier than the application date reported in the interference, and among interfering applicants, benefits dates tend to be closer together.

A, control patents are restricted to share a 6-digit technology class with both interfering patents, following Thompson and Fox-Kean (2005). In Panel B, control patents are restricted to share a 3-digit technology class with both interfering patents, and we use distance between first-named inventors, following Jaffe et al. (1993). Thus, while the observed (black line) distribution is the same in both panels, the counterfactual distribution is more geographically concentrated in Panel A.

We find that interferences are indeed significantly geographically localized, for either counterfactual. We take this as evidence that geographic proximity is related to the sharing of knowledge inputs.

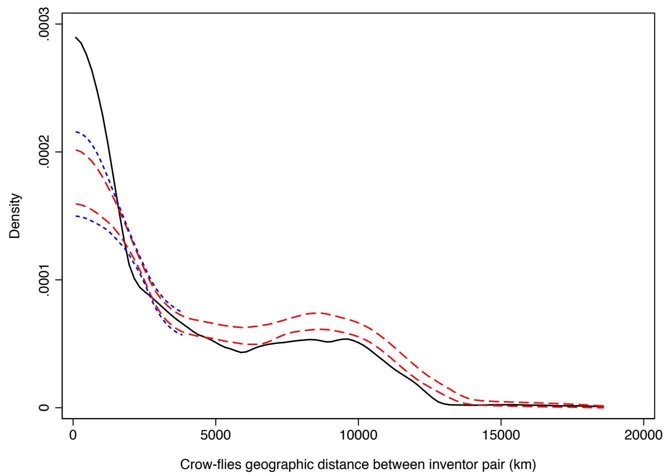
4.5.3 Social network proximity

Next, following Breschi and Lissoni (2005), we test whether the geographic localization of interferences might be related to closer social ties. Figure 2 compares the distribution of social network distances between pairs of interfering inventors to the distribution for similar, non-interfering pairs. 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 local confidence bands of the kernel-density for non-interfering, control pairs. The dotted blue lines show the global confidence bands of the kernel-density for non-interfering, control pairs. For each interfering and control pair we define social network distance as the length of the shortest distance path along our constructed inventor social network between any two co-inventors. Control patents are restricted to share a 6-digit technology class with both interfering patents. Thus, we find that interfering inventors tend to be in closer social proximity versus similar inventors not linked by common invention.

4.5.4 Comparison to citation-linked controls

Next, we compare the observed distribution of geographic proximity between interfering inventors to a particularly strong counterfactual: pairs of control

Panel A. Six-digit controls, minimum distance between inventors



Panel B. Three-digit controls, distance between first-named inventors

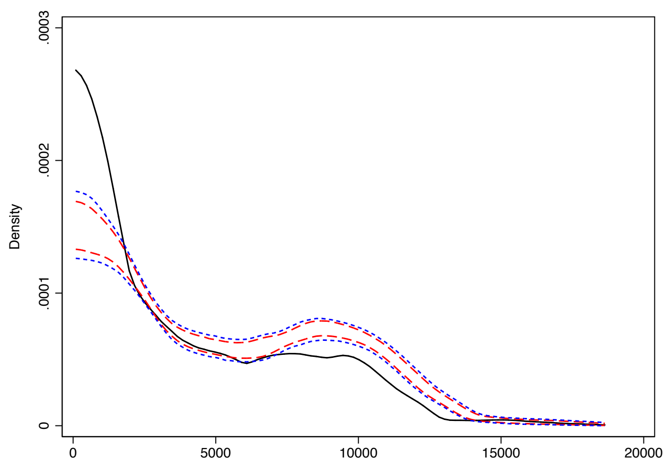


Figure 1: Distribution of geographic distances between pairs of inventors

Notes: This graph compares the distribution of geographic distances between pairs of interfering inventors to the distribution for similar, non-interfering pairs. 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 local confidence bands of the kernel-density for non-interfering, control pairs. The dotted blue lines show the global confidence bands of the kernel-density for non-interfering, control pairs. For each interfering and control pair we used the minimum pairwise distance among co-inventors. In Panel A, control patents are restricted to share a 6-digit technology class with both interfering patents. In Panel B, control patents are restricted to share a 3-digit technology class with both interfering patents.

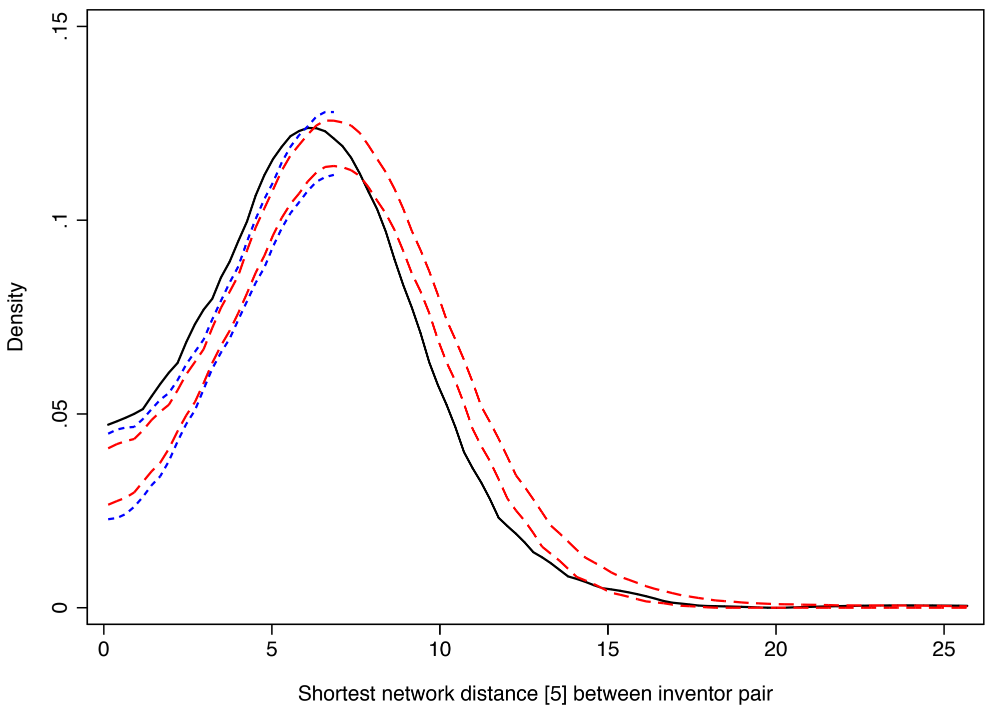


Figure 2: Distribution of network distances between pairs of inventors

Notes: This graph compares the distribution of social network distances between pairs of interfering inventors to the distribution for similar, non-interfering pairs. 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 local confidence bands of the kernel-density for non-interfering, control pairs. The dotted blue lines show the global confidence bands of the kernel-density for non-interfering, control pairs. For each interfering and control pair we define social network distance as the length of the shortest distance path along our constructed inventor social network between any two co-inventors. Control patents are restricted to share a 6-digit technology class with both interfering patents.

patents and interfering patents linked by citation. Thus, our "control pairs" in this section are the "treated" pairs in Jaffe et al. (1993). The motivation for this test is the idea that geographic proximity might be particularly important for tacit knowledge. If interfering inventors share such knowledge, then we might expect to see interferences more geographically concentrated versus even citation-linked control pairs.

Figure 3 shows this comparison. Even compared against the geography of citation-linked patents, interfering patents are significantly more concentrated.

5 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.

To do:

- By industry
- Robustness to decisions on priority
- Robustness to 100% overlapping claims
- Appendix

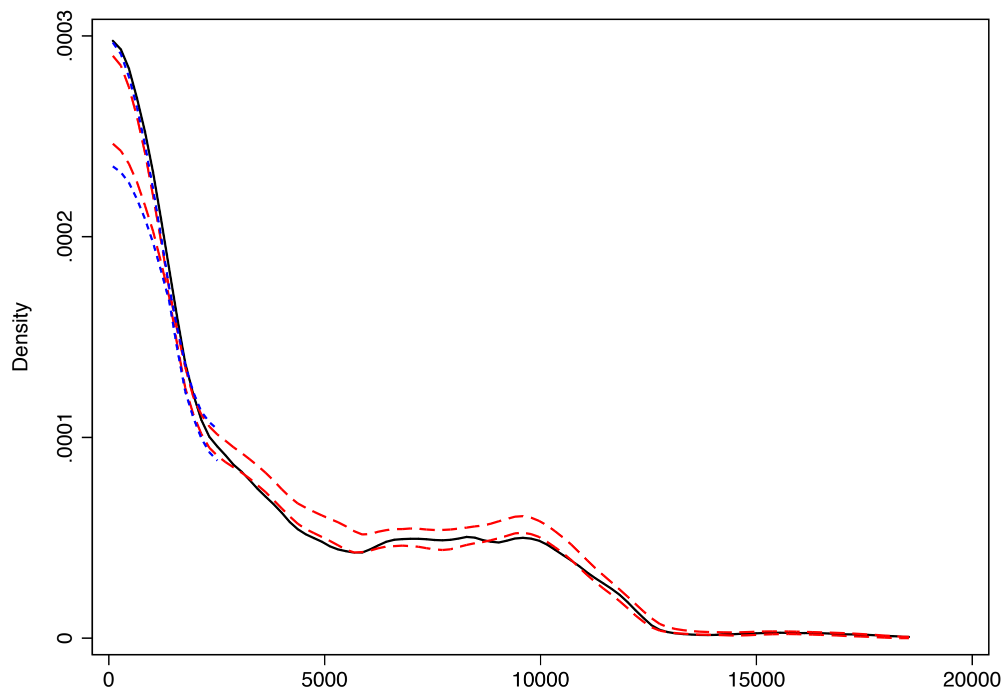


Figure 3: Distribution of geographic distances vs. citation-linked controls

Notes: This graph compares the distribution 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. 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 local confidence bands of the kernel-density for non-interfering, control pairs. The dotted blue lines show the global confidence bands of the kernel-density for non-interfering, control pairs. For each interfering and control pair we used the minimum pairwise distance among co-inventors. Control patents are restricted to backwards citations from one of the interfering patents.

6 Works cited

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

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., Jake K. Carr, Robert M. Hunt, and Tony E. Smith (2012). "The Agglomeration of R&D Labs," Federal Reserve Bank of Philadelphia working paper 12-22.

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt (2007). “Urban Density and the Rate of Invention,” Journal of Urban Economics 61: 389–419.

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.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. “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.

Kleid, Dennis G (2002). “Scientist and Patent Agent at Genentech,” an oral history conducted in 2001 and 2002 by Sally Smith Hughes for the Regional Oral History Office, The Bancroft Library, University of California, Berkeley.

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 UNF:5:RqsI3LsQEYLHkkg5jG/jRg= The Harvard Dataverse Network [Distributor] V5 [Version]

Lucas, Robert (1988). “On the Mechanics of Economic Development,” Journal of Monetary Economics 22: 3–42.

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. "Localized knowledge spillovers and patent citations: A distance-based approach," Review of Economics and Statistics, forthcoming.

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

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 Court of Appeals for the Federal Circuit. Decision 99-1604 in interference no. 102416, Hitzeman et al. v. Rutter et al.. March 21, 2001.

United States Patent and Trademark Office. "e-FOIA Reading Room," http://e-foia.uspto.gov/Foia/PTABReadingRoom.jsp, accessed XYZ XX, 2012.

United States Patent and Trademark Office. "Efiling for Patent Trial and Appeal Board," https://acts.uspto.gov/ifiling/PublicView.jsp, accessed XYZ XX, 2013.

Schmookler, Jacob (1966). Innovation and Economic Growth. Cambridge, MA: Harvard University Press.

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