Journal of Economic Geography **6** (2006) pp. 571–591
Advance Access published on 28 September 2006 doi:10.1093/jeg/lbl016

Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships

Ajay Agrawal, Iain Cockburn** and John McHale****

Abstract

We examine the role of social relationships in facilitating knowledge flows by estimating the flow premium captured by a mobile inventor's previous location. Once an inventor has moved, they are gone—but are they forgotten? We find that knowledge flows to an inventor's prior location are approximately 50% greater than if they had never lived there, suggesting that social relationships, not just physical proximity, are important for determining flow patterns. Furthermore, we find that a large portion of this social effect is mediated by institutional links; however, this is not the result of corporate knowledge management systems but rather of personal relationships formed through co-location within an institutional context that endure over time, space, and organizational boundaries. Moreover, we find the effect is nearly twice as large for knowledge flows across as compared to within fields, suggesting that co-location may substitute for communities of practice in determining flow patterns.

Keywords: knowledge flows, social relationships, labor mobility, patent citations
JEL classifications: F22, O31, R12, R23
Date submitted: 29 February 2004 **Date accepted:** 9 August 2006

1. Introduction

Prior empirical research has shown that knowledge flows are geographically localized (Jaffe et al., 1993, hereafter referred to as JTH). The measurement of this phenomenon is of particular importance because of the central role played by knowledge flows in growth theory. However, although economic theory offers explanations regarding why knowledge spillovers are important for economic growth (Romer, 1986, 1990) and why the localization of knowledge flows is important for regional advantage (Marshall, 1890; Krugman, 1991; Porter, 2000), the empirical literature is surprisingly thin on *why* knowledge flows may be localized in the first place.

Almeida and Kogut (1999) notably report results suggesting that interregional labor mobility may be a cause of knowledge localization. The authors' findings suggest regions such as Silicon Valley that experience higher than average levels of interfirm inventor mobility tend to also experience a greater degree of knowledge localization,

* Rotman School of Management, University of Toronto and NBER, Toronto, Ontario, Canada M5S 3E6. *email* <ajay.agrawal@rotman.utoronto.ca>
** Boston University and NBER.
*** Queen's University.

implying a direct relationship between labor mobility and knowledge flows.[1] However, the authors do not imply that it is necessarily the relocated inventors themselves that are citing their prior companies; rather inventors take knowledge with them and are able to share it with those to whom they are in close proximity. So why is it necessary for knowledge of patented inventions that are publicly disclosed to be physically carried by mobile engineers? Because, Almeida and Kogut say, an important component of the knowledge associated with patented inventions may be 'held tacitly by skilled engineers.'

In other words, practicing a patented invention may require complementary knowledge that is tacit (very costly or impossible to codify), or the complementary knowledge may be straightforward to codify but is not codified since this was not necessary for claiming priority over the invention. For example, certain failed experiments may be important for understanding how to modify an invention for alternative applications but are not codified since there is little incentive to do so (i.e. journals are usually not interested). So even though this knowledge might be quite easily codified, it is not. Agrawal (2006) provides a detailed example of a discovery in the area of robotics for which complementary knowledge was not codified and yet important for practicing the invention.

But is it necessary to be in close proximity in order to transfer tacit knowledge? The commonly described benefits of close spatial proximity for facilitating knowledge flows include (1) lower communication costs, (2) higher likelihood of chance meetings, and (3) higher likelihood of social relationships. In other words, individuals who are co-located are able to meet and exchange ideas at lower cost than those who are geographically separated. At the same time, individuals who are co-located are more likely to experience serendipitous meetings during which useful knowledge exchanges may occur. Finally, co-located individuals are more likely to develop social relationships, which may act as conduits for knowledge flows.

In this paper, we build on the prior research by exploring the particular role of social relationships in mediating knowledge flows. To what extent does the presence of a social relationship reduce the importance of spatial proximity in mediating knowledge flows? Not only will addressing this question shed light on an important determinant of non-local knowledge flow patterns, but it will indirectly help us better understand why many knowledge flows are in fact localized.

We remain agnostic to the socializing mechanism; we do not speculate on how individuals form relationships but rather consider co-location as a sort of social treatment that increases the probability of forming a social relationship. In fact, we do not even require social relationships to be direct. So, for example, if inventor A has a relationship with B, and B has a relationship with C, it is possible for B to facilitate a knowledge exchange between A and C since inventor B has a social relationship with both.

We examine knowledge flows between inventors and regions using patent citation data. One might expect social relationships to play a particularly minimal role in mediating knowledge flows in this context due to the strong incentives to disseminate scientific knowledge widely and rapidly (Dasgupta and David, 1987, 1994). In other words, inventors are highly motivated to fully articulate their new knowledge in the

1 In related work, Song et al. (2003) find evidence of firm 'learning-by-hiring,' which also suggests a relationship between labor mobility and knowledge flows.

claims of a patent or the text of a publication in order to claim priority over their invention. So, findings that suggest evidence for the importance of social relationships in mediating knowledge flows in the context of inventors may significantly understate the importance of this issue in more general settings.

Specifically, we collect data on 'movers' (inventors who have patented in one location and then patented somewhere else) and test whether knowledge generated in their new location flows disproportionately back to their prior location. We use this approach to identify the social relationship effect on knowledge flows; other benefits of co-location—lower communication costs and higher likelihood of chance meetings—are not relevant after previously co-located individuals are separated, but social relationships (may) persist.[2]

Thus, the first hypothesis we test is that knowledge flows go disproportionately to the inventor's *prior* location. We refer to this as the 'enduring social relationship hypothesis.' In Agrawal et al. (2003), we further motivate this hypothesis with a simple model of purposeful investments in social relationships between (potentially) mobile inventors. Our empirical findings reported in the present paper suggest strong support for this hypothesis; they are both statistically and economically significant. We estimate that knowledge flows are approximately 50% more likely to go to the inventor's prior location than if the inventor had never lived there.

Next, we turn to the implications for the social relationship hypothesis to explore within versus across-field knowledge flows. We argue that geography is likely to be less important in mediating social relationships between individuals in the same field since they have various alternative mechanisms through which to establish relationships. For example, individuals in the same community of practice (Brown and Duguid, 1991; Lave and Wenger, 1993) or invisible college (Crane, 1965, 1969) attend conferences and trade shows together, belong to common associations, and have other institutional settings in which to fraternize and share ideas.

Stated another way, spatial proximity will be more important for mediating social relationships between individuals from different fields. The sociological literature suggests some reasons for this. First, the work of Granovetter (1973) on the strength of weak ties points to how the extent of overlap between two individuals' networks is correlated with the strength of ties between them. Given that one's friends' friends are also likely to be their friends, the friend may actually provide them with little information they cannot get from the rest of their friendship network. In our case, friends can be considered as colleagues in the same field or invisible college. In contrast, an acquaintance to which they are only weakly tied (in our case, inventors not in the same invisible college but rather co-located) may provide them with truly novel information. More recently, Burt (1992) emphasized opportunities for value-creating brokerage that accrue to individuals who can fill 'structural holes' in networks—that is,

2 It is important to note, however, that the effective cost of communication is affected by the presence of a social relationship. In other words, while formerly co-located individuals who have formed a social relationship must bear the normal mechanical costs of communicating long distance (phone, flight tickets, etc.), the effort required is likely less than long distance communication between strangers. As such, the effective cost of communication is lowered due to the existence of a social relationship. We consider these lowered costs and the associated benefits with respect to knowledge flows a function of the social relationship.

relationships with people who do not have relationships with one another. Co-located inventors who do not belong to the same invisible college may create such brokerage opportunities.

Thus, the second hypothesis we test is that the proportionate increase in knowledge flows due to co-location is greater for flows *across* technology fields than within. Again, our results suggest strong support for this hypothesis. We estimate that the 'co-location premium' is nearly twice as large for knowledge flows across fields as compared to within.

The paper is structured as follows. In the next section, we outline our case control methodology for testing the two hypotheses described above. We describe our patent citation data in Section 3 and present our findings in Section 4. We discuss the implications of our results in Section 5 as well as offer directions for future research.

2. Empirical methodology

The two hypotheses we wish to test are both based on the geographic distribution of knowledge flows, yet knowledge flows are notoriously difficult to measure. Following the work of Adam Jaffe, Manuel Trajtenberg, and co-authors (see the collected papers in Jaffe and Trajtenberg, 2002), we use patent citation data as an indicator of knowledge flows between inventors.[3] We adapt the methodology of one of the seminal papers in this literature (JTH) in ways that allow us to examine the effects of inventor mobility on the geography of knowledge flows.[4]

The essence of the JTH methodology is the comparison of citing patents with control patents in terms of the frequency with which each is located in the same region as the focal patent. The finding of a disproportionate number of co-located citations relative to co-located control patents is interpreted as evidence of localized knowledge flows.

3 Patent citations are not straightforward to interpret in terms of communication between inventors, and the signal-to-noise ratio for this measure is therefore likely to be rather low. Patents cite other patents as 'prior art,' with citations serving to delineate the property rights conferred. Some citations are supplied by the applicant, others by the patent examiner, and some patents may be cited more frequently than others because they are more salient in terms of satisfying legal definitions of prior art rather than because they have greater technological significance. Cockburn et al. (2002) report, for example, that some examiners have 'favorite' patents that they cite preferentially because they 'teach the art' particularly well. Nonetheless, Jaffe et al. (2002) surveyed cited and citing inventors to explore the 'meaning of patent citations' and found that approximately one-quarter of the survey responses corresponded to a 'fairly clear spillover,' approximately one-half indicated no spillover, and the remaining quarter indicate some possibility of a spillover. Based on their survey data, the authors conclude: 'We believe that these results are consistent with the notion that citations are a noisy signal of the presence of spillovers. This implies that aggregate citation flows can be used as proxies for knowledge-spillover intensity, for example, between categories of organizations or between geographic regions' (p. 400). Furthermore, the authors draw the following general conclusion regarding the relationship between citations and knowledge flows: 'Overall, the results confirm that citations can be interpreted as providing a (noisy) signal of spillovers' (p. 394). Moreover, the fact that only a fraction of citations are driven by inventor contact is reflected in our finding of a small albeit statistically significant effect. Finally, as noted in the text, we allow for indirect social relationships that are not captured in the Jaffe survey. So, for example, if inventor A has a relationship with B and B has a relationship with C, it is possible for B to facilitate a knowledge exchange between A and C since B has a social relationship with both.

4 Thompson and Fox-Kean (2005) recently proposed a method for enhancing the selection of control patents by matching on a primary and secondary six-digit classification rather than the three-digit primary classification of the citing patent.

The reason for using controls is that patent citations will tend to be co-located with the focal inventions even in the absence of knowledge flows when inventive activity in particular technological areas is clustered geographically.[5] Thus, the localization effect is identified as the extent to which citations are from inventors who are co-located with an inventor of the focal patent over and above what we would expect given the geographic concentration of inventive activity in the particular technological area of the citing patent.

More formally, we define the probability of co-location in our sample of control patents as the unconditional probability of co-location, $P(\text{Co-location})$, and the probability of co-location given that an actual citation has occurred as the conditional probability of co-location, $P(\text{Co-location}|\text{Citation})$. Our basic hypothesis test is that the difference between the conditional and unconditional probabilities—i.e. the co-location premium—is positive and statistically significant. In economic terms, however, it will often make more sense to think about how inventor co-location affects the probability of a citation rather than how the occurrence of a citation affects the probability of the inventors being co-located. In other words, the interesting causal relationship is from co-location to the likelihood of a knowledge spillover (as proxied by a citation). Of course, the two probabilities are related by Bayes Rule.

$$\frac{P(\text{Citation}|\text{Co-location})}{P(\text{Citation})} = \frac{P(\text{Co-location}|\text{Citation})}{P(\text{Co-location})}. \tag{1}$$

Subtracting 1 from both sides, we see that the proportionate increase in the probability of a citation conditional on co-location is equal to the proportionate increase in the probability of co-location conditional on a citation. Thus, in presenting our results, we also report the ratio of $[P(\text{Co-location}|\text{Citation}) - P(\text{Co-location})]$ to $P(\text{Co-location})$. This ratio measures the proportionate increase in the probability of a citation due to co-location.

As in JTH, a simple measurement of location matches would not account for any geographic clustering of innovative activity within particular technological areas. For example, an inventor on a patent for a particular type of medical device might be located in Boston, and the patent might receive a large fraction of citations from patents that include at least one inventor located in Boston. This might reflect knowledge flows through social relationships, or it could simply reflect the large fraction of overall patenting for medical devices that occurs in Boston.

We use the following procedure to construct the set of control patents. A control patent is selected for each observation that matches the citing patent on the following dimensions: (1) application year and (2) technology classification. Having generated the set of patents with the same application year and same original three-digit US classification as the citing patent, we identify the patent in the set that has the closest grant date to the citing patent. Next, we confirm that the control patent does not cite the focal patent. If it does, we remove the patent from the set of potential control

5 The clustering itself may be due to the localization of knowledge flows but also may be due to other local factors, such as thicker factor markets. So focusing on knowledge flows that are concentrated more than the innovative activity in that particular field may be considered a conservative approach.

patents and select the next best control patent. Finally, if there are no patents that match both the citing patent's application year and original classification without citing the focal patent itself, then the observation (focal patent) is removed from the dataset.

The unit of analysis in JTH is a citation to an originating *patent*. Given our focus on knowledge flows between individual inventors, our unit of analysis is a citation to an *inventor* on an originating patent—what we call an inventor–patent-citation. Thus, a single patent that has two inventors and is cited by five subsequent patents will generate 10 unique observations.[6] Each observation is assigned to a location (MSA) based on the city and state information associated with the home address of the inventor as reported on the front page of the patent.

This unit of analysis allows us to consider knowledge flows between people rather than between patents. Accordingly, we are able to observe the pattern of knowledge flows that results when individuals move locations, which is the basis of our method for isolating social relationship effects from communication costs or chance meetings. To do this, we identify 'movers' in our data.

Movers are individuals who have patented prior to their focal 1990 patent while living in a North American MSA that is different from the MSA they were in at the time of their 1990 patent. Specifically, we identify 'movers' in these data by examining the inventor names on all our 1990 patents. We then identify all patents applied for in prior years that have matching inventor names.[7]

However, matching inventors purely on their names introduces the risk of false negative errors (inventors may use multiple spelling permutations of their name such that we miss actual movers) and false positive errors (different inventors may have the same name such that we flag someone as a mover who is not) (Trajtenberg, 2004). We do not address false negative errors and thus our sample is a conservative estimate of the overall levels of inventor migration. However, since we do not expect the likelihood of recording different name spellings across multiple patents to be correlated with co-location citation propensities, we do not expect this measurement error to bias our main result.

To minimize false positive errors, we add the sampling restriction that the inventor's pre-1990 patent must be in a similar technology field as their 1990 patent. Therefore, we impose the restriction that the 1990 patent's major three-digit classification must either match the prior patent's own major three-digit classification or be included in the prior patent's set of three-digit cross-classifications. Thus, we eliminate the problem of falsely identifying two inventors with the same name who are working in different fields and in different locations as the same person and hence a mover.[8] However, it is important to note that we may still experience measurement error in our process for identifying movers. To the extent that we falsely identify movers (two or more individuals who have the same name and patent in the same field but live in different cities), our findings will be biased against our hypothesis.

6 Such a patent would only generate five observations using the JTH method since their unit of analysis is the patent-citation rather than the inventor–patent-citation.

7 We match on full names, including the middle initial if present.

8 Also, if the inventor patented in more than one location prior to their 1990 location, we select the patent with the application date that is closest in time to that of the focal patent (1990).

We test our first hypothesis by comparing the proportion of citing patents that are from the mover's prior location with the proportion of control patents that are from the mover's prior location. We report results from tests on the equality of these proportions using large-sample statistics.[9]

We test our second hypothesis that the proportionate increase in knowledge flows due to co-location is greater across technology fields than within by splitting the sample into two groups. The first group contains those observations in which the citing patent is from the same field as the focal patent and the second group contains those in which it is not. We compare the co-location premiums (the difference in proportion of citing patent location matches with control patent location matches) across the two groups by taking differences-in-differences.

We employ two schemes for classifying patents as being from the same field. First, we classify those citing patents with the same two-digit NBER patent subcategory classification as the focal patent as being from the same field.[10] Next, we classify those citing patents with the same three-digit US patent classification as the focal patent as being from the same field. The two methods produce similar results;[11] only the NBER subcategory results are reported here.

Finally, it is important to recall that we focus on citations to an inventor's *new* inventions. That is, we examine the degree to which citations to an invention made in an individual's new location (after they moved) come disproportionately from the inventor's prior location. To the extent that new inventions are simply extensions of previous inventions that the inventor developed in their prior location, it may be the case that formerly co-located engineers learned about an earlier invention and thus better understood the new invention (and were thus more likely to cite it). Therefore, we must make the restrictive assumption that new patented inventions are sufficiently unique that exposure to prior inventions is not solely driving the propensity for former neighbors to disproportionately cite the inventor in their new location.[12]

3. Data

We use the 'front page' bibliographic data for patents published by the United States Patent and Trademark Office (USPTO) as the basis for most of the empirical work. These data contain the application date and issue date of each patent, the names and locations of inventor(s), a technology classification, and a list of other patents cited. We augment these data with the NBER Patent-Citations data file for additional fields, including the one-digit technology category code, the two-digit subcategory code, and the assignee code.

9 Specifically, we report the test statistic generated by the 'prtest' function from Stata 7.0.

10 There are 35 two-digit NBER subcategories. The categories are generally designed to reflect industry classifications (like SIC classifications) as opposed to technology classifications.

11 We compare classification schemes in the context of the 'Drugs and Medical' category to offer some sense of the relationship between these two. This NBER-defined *category* encompasses four *subcategories* and 14 three-digit *US classifications*. For example, the four subcategories include drugs, surgery and medical instruments, biotechnology, and miscellaneous. One of these subcategories, drugs, encompasses two US classifications, 424 and 514, which are both described as 'drug, bio-affecting, and body treating compositions.'

12 We thank an anonymous referee for bringing this caveat to our attention.

We begin with the full set of issued patents that have their application year as 1990.[13] There are 108,672 such patents. From these, we select the set of patents that are from North America.[14] There are 60,974 such patents. We then discard all patents that have not received any citations, since our study is based on examining citations as a proxy for knowledge flows. We do not believe this elimination results in selection bias since we are interested in measures that are conditional on there being citations. Consequently, approximately 8.7% of the remaining patents are discarded, leaving 55,664 as the set of 'originating patents' that form the basis of the empirical analysis.

Each of the originating patents has an average of approximately 10.2 citations, resulting in 568,960 unique patent citations. A small fraction of these observations are removed because their citing patents do not map to an MSA or because our process for generating control patents is not able to find an adequate control for the citing patent. This process reduces the number of observations to 564,590. Next, we discard the 11.6% of observations for which the citing patent is a self-citation by one or more of the inventors, since a self-citation does not reflect a knowledge flow from one individual to another.[15] This leaves us with 499,341 observations. Finally, we 'unbundle' individual inventors, of which there are an average of approximately two per patent, resulting in a final sample size of 992,362 observations. For the part of the analysis that is based only on movers, we select those observations associated with individuals who had patented prior to 1990 in a different MSA. This results in 59,734 observations, which represents approximately 6% of the full sample.

We follow an identical procedure for generating the 1989 dataset. There are slightly fewer North American patents in 1989 (56,896 rather than 60,974). Ultimately, we generate 938,419 observations for the 1989 cohort.

Every observation is assigned to an 'originating location' based on the home address of the inventor. Inventors are assigned to an MSA based on their city and state information.[16] There are 268 US MSAs and consolidated

13 We replicate the entire study with 1989 patents, which is a completely distinct set from 1990, and also report these results throughout the paper. In all cases, the results are similar across the two years. In addition, we conduct some analyses with 1975 and 1980 data in order to offer a direct comparison with the JTH study, which uses data from those years. These comparisons are reported in Agrawal et al. (2003). Our data confirm their earlier findings.

14 We use the geographic assignment procedure developed by JTH to determine whether patents are from North America. This procedure works as follows. Where there is a single inventor, the patent is assigned to the location of that inventor. Where there are multiple inventors, the patent is assigned to the location of the majority of inventors. In other words, if there are two inventors from Boston and one inventor from Paris, the patent is assigned to Boston. If there is a tie across inventor locations (e.g. one inventor in Boston and the other in Paris), the patent is randomly assigned to one of these locations. Finally, it is important to note that some North American inventors are located in regions that are not mapped to an MSA. In these cases, we assign the patent to a 'phantom MSA.' Phantom MSAs are created for each US state and Canadian province.

15 We consider assignee name matches or inventor name matches as self-cites. This is perhaps a stricter definition than often used in citation-based empirical research, which often only considers assignee name matches as self-citations. Since we are particularly concerned with 'movers,' we want to eliminate the possibility of an individual citing their own prior work while at a new firm and thus filing under a new assignee name, since this does not represent a knowledge flow from one individual to another.

16 City and country information is used for assigning Canadian inventors to a CMA.

Table 1. 10 Largest MSAs in terms of number of observations in dataset

MSA	Number of observations with inventor of originating patent from specified MSA	Percentage of total observations
San Francisco–Oakland–San Jose, CA	99,414	10.0
New York–Northern NJ–Long Island, NY NJ CT	98,833	10.0
Boston–Worcester–Lawrence, MA NH ME CT	57,503	5.8
Los Angeles–Riverside–Orange County, CA	53,186	5.4
Chicago–Gary–Kenosha, IL IN WI	43,238	4.4
Minneapolis–St Paul, MN WI	29,059	2.9
Philadelphia–Wilmington–Atlantic City, PA NJ DE MD	28,000	2.8
Detroit–Ann Arbor–Flint, MI	24,240	2.4
Dallas–Fort Worth, TX	22,681	2.3
Rochester, NY	20,714	2.1
Total for 10 largest MSAs	476,868	48.1

metropolitan statistical areas (CMSAs) and 25 Canadian census metropolitan areas (CMAs)—hereinafter collectively referred to as 'MSAs.'[17,18] We have also created 63 'phantom MSAs' for individuals located in one of the 50 states or 13 provinces or territories that are in cities not assigned to one of the Census Bureau-defined MSAs.

To this end, our observations are not distributed evenly across MSAs. In fact, the 10 largest MSAs, in terms of number of observations where the inventor is located in that MSA, account for almost half the sample. This is illustrated in Table 1. As described in the methods section, we deal with the heavily skewed nature of these data by constructing a set of control patents that is intended to account for the uneven distribution of innovative activity across geographic space.

17 While MSAs and CMAs are similar in spirit, they are defined slightly differently. The Canadian criterion requires that the urban core have a population of at least 100,000 for a metropolitan area to exist. In contrast, for the period 1990–2000, the United States had two criteria to determine whether or not a metropolitan area existed: (1) where there is either a city of 50,000 or more inhabitants or (2) where there is a Census Bureau-defined urban area, i.e. a population of at least 50,000 and a total metropolitan population of at least 100,000 (75,000 in New England). Thus, the Canadian approach is the more restrictive of the two.

18 While not perfect, the MSA does satisfy the key criterion of reflecting regions that are socially and economically integrated. In other words, unlike politically designated geographic regions such as states, MSAs reflect regions of concentrated employment, such as city centers, and the adjacent regions where commuting workers live. For example, the San Francisco MSA includes San Francisco, Oakland, and San Jose as well as the nearby smaller cities and towns that feed into this economic region. Certainly, these geographic units are noisy in that people may live in one MSA but work in another. In addition, some MSAs are very large (the New York–New Jersey–Long Island MSA is the largest) such that the probability of individuals who are co-located in the same MSA having a social relationship may be extremely small. To this end, we are planning future research that examines this phenomenon using smaller units of analysis, such as zip codes. Still, it is reasonable to assert that individuals working in one MSA are more likely to establish social relationships with others co-located in the same MSA than otherwise; since we are focused on these probabilities, we feel this unit of analysis is suitable.

Table 2. Comparing movers to general population: flows to their current location. Percent of citing and control patents in the same MSA as the originating patent.

	1990		1989	
	Full sample	Movers	Full sample	Movers
% Citing matching	8.6	8.4	8.4	8.1
% Controls matching	5.5	5.4	5.2	5.1
Co-location premium	3.1	3.0	3.2	3.0
z-statistic	85.15	20.64	86.08	20.75
Co-location premium/% controls matching	0.56	0.56	0.62	0.58
n	992,362	59,734	938,419	57,878

Finally, since much of our focus is on movers, it is important to note that our sample size drops significantly as we move from our full inventor sample to our mover sample. Though we still have almost 60,000 observations on movers, this is only slightly more than 6% of the full sample of inventor–patent-citation observations, raising the possibility that selection bias is affecting our results. To investigate whether movers are systematically different than the original sample population of inventors, we compare the two samples in terms of flows to the inventor's current location as well as other patenting characteristics. Table 2 presents the results from comparing these two groups in terms of co-location premiums. Movers' flows to their current location do not appear to be measurably different than those associated with the full sample.

We also compare movers with the general population along other dimensions in Table 3. We see the two samples are similar in terms of 'impact' as measured by the average number of citations received, in terms of distribution across types of assignees, and in terms of distribution across technical categories, although movers seem to be less concentrated in computers/communications and more concentrated in chemical and drugs/medical than the full sample.[19] Though we find no reason to believe that systematic differences between movers and non-movers are driving our results, we intend to look at differences in these two groups in greater detail in future research.[20]

19 We would like to compare the productivity of movers with that of non-movers. However, this would involve constructing the overall research profiles of the individuals in both populations since doing so would require controlling for years since graduation, etc. Constructing such profiles is beyond the scope of this paper since our main findings are not predicated on a comparison between the two types of inventors. However, as noted in the text, we are able to compare the 'impact' of the average mover invention with that of the average non-mover invention by comparing the average number of forward cites received. In Table 3, we see that the two types of inventors receive approximately the same amount of forward citations, on average.

20 We find that approximately 16% of the movers in our 1990 data moved back to their prior location (at least temporarily) sometime between 1991 and 2004. We plan to investigate the effect of these temporary movers in future research, particularly in the context of international labor mobility and national diasporas.

Table 3. Comparing movers with the general population: patenting characteristics (1990)

	General population $n = 990{,}524$*	Movers $n = 62{,}817$
Average number of citations received	21.8	20.5
Assignee code		
Unassigned	13.5%	9.4%
Assigned to a US non-government org.	81.8%	87.0%
Assigned to a non-US, non-government org.	2.1%	1.7%
Assigned to a US individual	1.0%	0.7%
Assigned to a non-US individual	0.0%	0.0%
Assigned to the US (Federal) Government	1.5%	1.0%
Assigned to a non-US government	0.1%	0.1%
Technological category		
Chemical	14.3%	19.8%
Computers and communications	25.3%	14.4%
Drugs and medical	15.4%	20.9%
Electrical and electronic	16.4%	18.4%
Mechanical	12.4%	12.0%
Other	16.2%	14.7%

*This dataset was generated by merging our database with the patent data available on the NBER website.

4. Results

4.1. H1: The enduring social relationships hypothesis

Here, we focus our attention on the premium captured by the inventor's previous location. Once an inventor has moved, they are gone—but are they forgotten? While close spatial proximity between inventors proxies for a number of factors that may influence knowledge flows, including lowered communication costs and higher probability of chance meetings, we hypothesize that it is also a proxy for higher probability of establishing a social relationship. As such, even after individuals are separated, a disproportionate level of knowledge may flow to the inventor's prior location since relationships can persist despite separation.

Our results, presented in Table 4, support this hypothesis. The frequency of citing patents matching the inventor's previous location is significantly greater than the frequency of control patents matching the inventor's previous location. The premiums (differences in proportions) are statistically significant with z-statistics of approximately 14 for both the 1990 and 1989 data cohorts. They are also economically significant; for the 1990 data, citing patents are 50% more likely than control patents to be located in the inventor's previous location.[21] Using Equation (1), this finding can be interpreted as indicating that prior co-location increases the probability of a citation by 50%. Stated another way, on average, 50% more citations are expected to come from the prior location than if the inventor had never lived there.

Also, it is interesting to note that the proportion of control location matches is substantially larger in the current location than in the prior location (5.1% compared to

21 $1.7/3.4 = 0.50$.

Table 4. Spillover premiums associated with movers: current versus prior locations. Percent of citing/control patents in current/previous MSA.

	1990		1989	
	Matching with current location	Matching with previous location	Matching with current location	Matching with previous location
% Citing matching	8.4	5.1	8.1	5.2
% Controls matching	5.4	3.4	5.1	3.6
Co-location premium	3.0	1.7	3.0	1.6
z-statistic	20.64	14.45	20.75	13.78
Co-location premium/ % controls matching	0.56	0.50	0.58	0.44
n	59,734	59,734	57,878	57,878

3.6% in 1989, and 5.4% compared to 3.4% in 1990). This suggests that, on average, movers relocate to regions where there is more activity in their technology area. While this may not seem surprising, the magnitude of the difference in levels of activity between prior and current locations is quite large. For example, on average, a control 1990 patent is 59% more likely to be found in the inventor's current location than in their prior location.[22]

Although we remain agnostic to the socializing mechanism, our data do offer some initial insights into the nature of these apparently persistent relationships. Table 5 illustrates the total number of observations where the citing (control) patent is from the inventor's prior location. The data in this table allow us to decompose the 'prior co-location premium' in terms of institutional relationships (inventor of citing or control patent worked for focal patent inventor's prior organization) and collaboration relationships (former co-inventors).

Examining the 1990 data, we see that 18% of the premium is due to individuals who were both former co-inventors and who were at one time associated with the inventor's prior organization.[23] A further 62% of the premium is due to individuals who were never co-inventors but who were at one time associated with the inventor's prior organization. Finally, the remaining 21% of the premium is due neither to a co-inventor link nor an organizational link. These individuals may be linked by a common third party, by other social group affiliations, as neighbors, or otherwise.[24]

Summing the first two fractions described above, a striking 80% of the premium is due to individuals who were at one time associated with the focal inventor's prior

22 $(5.4-3.4)/3.4 = 0.588$

23 We calculate this by taking the ratio of the difference between the number of citing versus control observations conditioned on the specified relationship characteristics (213–37) and the difference between the unconditioned number of citing versus control observations (i.e. the 'premium') (3034–2028); the ratio is thus $176/1006 = 17.5\%$.

24 Singh (2005) presents evidence that indirect social relationships, as measured by common third party co-inventors, do, in fact, influence knowledge flow patterns.

Table 5. Relationship between inventor of focal patent and inventor of citing (control) patent from prior location

	At least one citing (control) inventor from prior firm	No citing (control) inventors from prior firm
1990 data*		
At least one citing (control) inventor a co-inventor on a previous patent	213 (37)	0 (0)
No citing (control) inventors a co-inventor on any previous patents	834 (210)	1987 (1781)
1989 data**		
At least one citing (control) inventor a co-inventor on a previous patent	271 (35)	1 (0)
No citing (control) inventors a co-inventor on any previous patents	808 (260)	1939 (1764)

*Total number of citing (control) observations from prior location: 3034 (2028).
**Total number of citing (control) observations from prior location: 3019 (2059).

organization.[25] Furthermore, it is important to recall that we discard observations that involve inventor or assignee self-citations. Therefore, we know that none of this portion of the premium is a result of movers who moved locations within the same organization and received citations from former colleagues who work for the same organization but simply in a different location. This is particularly important since 50% of our sample observations (1990) are associated with intra-organizational moves.

Therefore, the disproportionate number of citations from the focal inventor's prior location is largely due to knowledge flows returning to individuals who were at one time also associated with the inventor's prior organization—which is a *different* organization than the one with which the inventor is now associated.[26] Moreover, the data reveal that 69% of the citing inventors who were at one time associated with the focal inventor's prior organization were no longer so at the time they filed their citing patent.

In other words, the identified effect is not the result of institutional knowledge management systems designed explicitly to direct flows across different locations within the same organization, but rather it is the result of personal relationships, formed within an institutional context, that endure over time, space, and organizational boundaries. To be clear, we still remain open to the notion that the focal and citing inventors may not have a *direct* personal relationship but that their temporary common institutional affiliation in the same geographical location may facilitate broad social networks that in turn facilitate subsequent knowledge flows.

25 The 1989 data reveal a very similar decomposition; 81.7% of the premium is due to individuals who were at one time associated with the focal inventor's prior organization.

26 Conversely, it could be that the focal inventor moved out of the original MSA but stayed within the organization and the citing inventor moved out of the organization but stayed within the original MSA.

4.2. Causal interpretation: social relationships versus distance

Our hypothesis is that knowledge flows by movers go disproportionately to their prior locations, relative to the case where they had never lived in that prior location. Moreover, we hypothesize that the effect is *causal* in the sense that the inventor's social relationships are specific to their prior location, and that the enduring element of these social relationships facilitates subsequent communication between the inventor in question and other inventors in their prior location.

The findings reported above are certainly consistent with this hypothesis. Citations occur disproportionately from the inventor's prior location when compared with the geographic distribution of a well-specified set of control citations. But we could observe such disproportionate citing to prior locations without the relationship being causal. Our greatest concern is that there is some omitted variable that affects both labor flows and knowledge flows. The most likely candidate for such an omitted variable is distance.

For example, suppose we look at the geographic distribution of citations to a 1990 patent with a New York inventor address, and that this inventor is observed to have applied for a patent in 1985 from a Boston address. Furthermore, we observe that a disproportionate number of cites to the 1990 patent occur from Boston, the inventor's 1985 location. We are tempted to view this as evidence for our enduring social relationships hypothesis, in the sense that Boston inventors are disproportionately citing the 1990 New York inventor because they continue to communicate with that inventor through the relationships and networks developed when the inventor was living and working in Boston.

However, an alternative explanation for the disproportionate cites is that Boston is relatively close to New York, so that New York inventors interact more regularly with Boston inventors than they do with inventors who live further away. If it is also true that, conditional on having moved from somewhere, a 1990 inventor is more likely to have moved from somewhere close (Boston to New York in our example), we will observe disproportionate cites to prior locations in our data even without the causal effect we hypothesize being present. Put differently, distance affects both the probability of citation in a given location and the probability of that location being the inventor's prior location; our results may therefore be confounding the effect of distance with any effect of enduring social relationships.

Therefore, we attempt here to identify the causal effect. The identifying assumption is that distance is the omitted variable, and that the social relationships-based causal effect is invariant to how far the inventor has moved. This is clearly a strong assumption. It is conceivable that distance affects the durability of social relationships, since it may be costlier for former neighbors to maintain social relationships the further apart from each other they live.[27] If distance does affect the durability of social relationships, however, this identification strategy will bias against finding a causal effect, and we thus see our results as providing a lower bound for this effect. To implement this strategy, we choose a matched location for each mover that is approximately the same distance from the inventor's final location as the distance between the inventor's prior location and

27 It would also be interesting to check for any influence of locations being in the same time zone, frequency of non-stop flights, or other factors affecting the cost of maintaining relationships.

final location. The difference between the premiums for the actual and matched prior locations is then identified as the causal effect.

Consider an originating patent filed in 1990 by an inventor living in Austin, TX. Suppose that the likelihood of citation to the 1990 patent by any given location is, all else equal, negatively related to the distance of that location from Austin. Thus, for example, the likelihood of a citation from Denver, CO, is greater than the likelihood of a citation from Portland, OR, since Denver is closer than Portland to Austin. Suppose further that given an inventor has moved to Austin, the likelihood that the inventor came from any particular location is, all else equal, also negatively related to the distance of that location from Austin. It follows that citations will disproportionately occur from the locations from which inventors moved, even when there is no social relationships-based causal effect.

An obvious way to isolate the causal effect is to find control locations that match the inventor's previous location in terms of distance from their 1990 location. We use a two-step procedure for identifying 'matching MSAs.' First, we measure the distance between the inventor's prior location and their 1990 location.[28] Then, we identify all other MSAs that are the same distance from the inventor's 1990 location, plus or minus 100 miles.[29] From this set of MSAs, we select the MSA that is closest to the inventor's previous location in terms of number of patents.[30,31] Thus, we select a control MSA that is similar to the inventor's previous MSA in terms of both its distance from the 1990 location and its level of technological activity. Figure 1 illustrates an example. In this case, the inventor moved from Portland, OR, to Austin, TX. A band is created to identify all other MSAs that are approximately the same distance as Portland is from Austin. Over 20 such MSAs are identified. Portland is a mid-sized MSA in terms of inventive activity. From the set of MSAs that satisfy the distance from Austin criteria, Seattle, WA, has the closest number of 1990 patents and is therefore selected as the control MSA for that observation.[32]

We then compute the difference-in-differences between citing and control patents for prior and matching MSAs. The results are shown in Table 6 for both 1990 and 1989 data. Again, we focus only on the 1990 results, as the two sets of results are almost identical. The prior location premium in the actual prior locations (1.6%) is substantially higher than it is in the matched prior locations (0.6%). The difference in the prior

28 For this step, we measure the distance from city to city.

29 For this step, we measure the distance from MSA to MSA. The distance between MSAs is measured between the largest cities within each of the MSAs.

30 For the purposes of comparing patenting activity across MSAs, we use patents with 1990 (or 1989) application dates and assign patents to MSAs by inventor location, using the 'majority rules' location determination method, as described earlier.

31 Not all candidates within this set are likely to be equally good matches for the actual prior location. Given the tendency for inventive activity to cluster, one concern is that the actual prior locations are more likely to be major metropolitan areas than randomly chosen locations from the candidate set. Although this concern is partly allayed if we compare the actual citation pattern with the control citation pattern for both the actual and the matching prior location—that is, use a difference-in-difference estimation approach—we are still concerned that matching prior locations are systematically different from the actual prior locations. For this reason, we choose the location that comes closest to the actual prior location in terms of overall 1990 patent applications.

32 Just as we only consider movers who have moved from one North American MSA to another, we restrict our search for matching MSAs to North America.

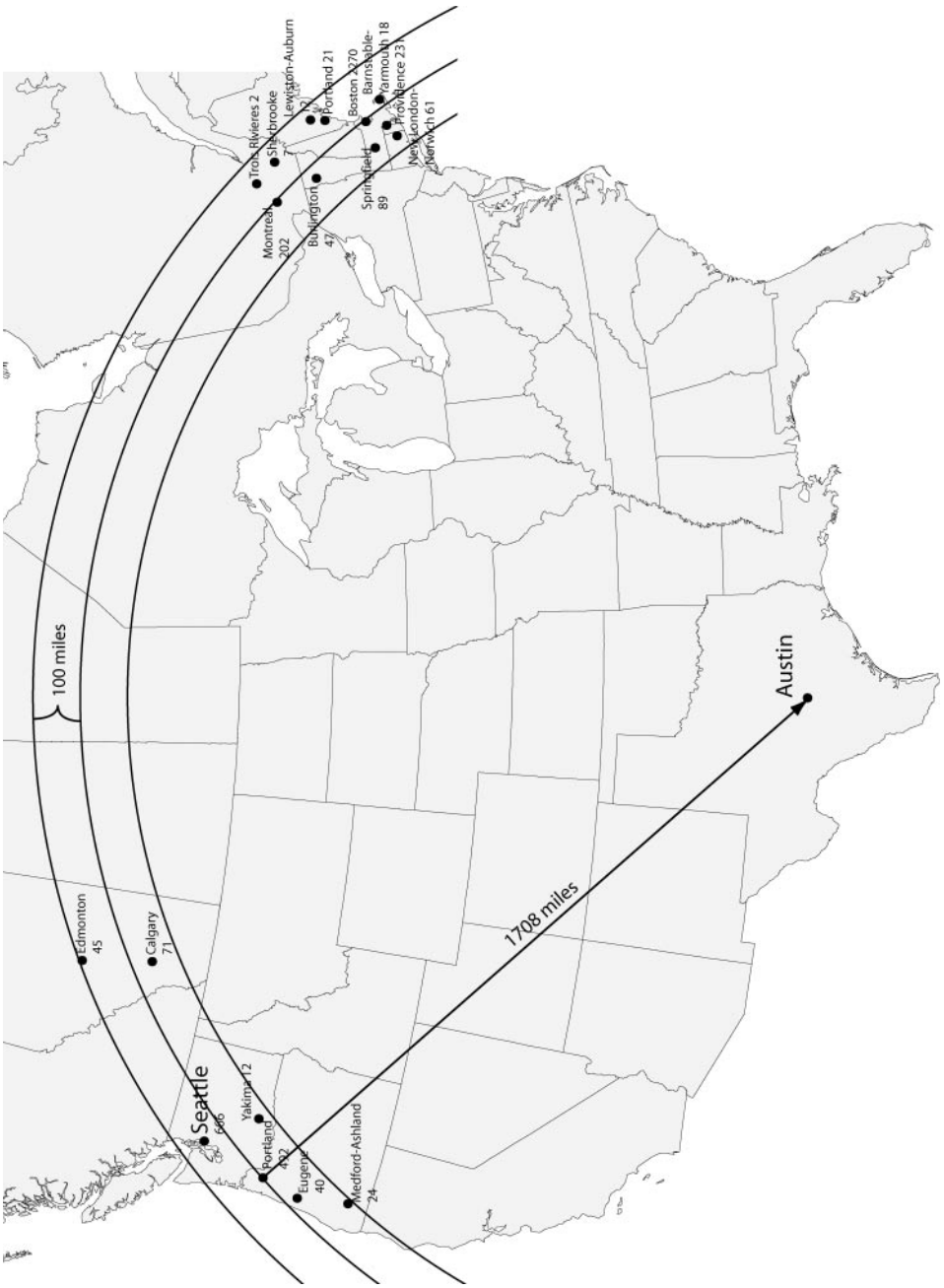

Figure 1. Selection of matching MSAs.

Table 6. Spillover premiums: comparison of actual prior locations with prior locations matched on distance

	1990	1989
Actual prior locations		
% Citing matching	4.1	4.2
% Controls matching	2.6	2.7
Prior co-location premium	1.6	1.5
z-statistic ($P > z$)	12.97 (0.00)	11.98 (0.00)
Prior co-location premium/% controls matching	0.61	0.55
n	44592	43100
Matched prior locations		
% Citing matching	1.5	1.3
% Controls matching	0.9	0.8
Prior co-location premium	0.6	0.5
z-statistic ($P > z$)	8.30 (0.00)	6.69 (0.00)
Prior co-location premium/% controls matching	0.67	0.62
n	44592	43100
Difference in differences		
Difference in prior co-location premium between actual and matched prior locations	0.9	1.0
t-statistic	7.22	7.62

location premium for actual and matched samples—i.e. the difference-in-differences—is clearly statistically significant with a z-statistic of almost 14.

4.3. H2: Knowledge transfer across technology fields

Flows of knowledge across technology fields (or between communities of practice) may rely more strongly on social relationships that are facilitated by co-location than do flows within a technology field. Because inventors have fewer alternative means for accessing new knowledge between different technology fields than they do for accessing new knowledge within their own field, the marginal benefit of geographically based social relationships may be higher for flows across fields than within.

The results are presented in Tables 7 and 8 for both the current and prior co-location premiums. Table 7 examines the difference in the current co-location premium for the full sample of inventors; Table 8 examines the difference in both the current and prior co-location premiums for the subsample of movers.[33]

Interestingly, the percentage of citation location matches is very similar for the two types of flows (cross-field compared to within-field) for both current and prior locations. Given the aforementioned fact of geographical concentration of activity by technological field, however, it is not surprising that the percentage of location matches for the controls is greater for the within-field flows in both cases.[34] It follows that the

33 Since we show results for current and prior co-location premiums in Table 8, we only present 1990 data for ease of display. However, almost identical results are obtained for the 1989 data.

34 Recall that the control patents are chosen to maximize the likelihood that they have the same technology classification as the actual citing patent. The reason that we expect to see a large number of location matches for the control patents is that technological activity is geographically concentrated by field. If,

Table 7. Spillover premiums (full sample): co-location premiums for cross-field versus within-field flows

	1990	1989
Cross-field flows		
% Citing matching	8.8	8.6
% Controls matching	5.0	4.7
Co-location premium	3.8	3.8
Co-location premium/% controls matching	0.76	0.79
z-statistic	80.25	79.33
n	577,851	528,047
Within-field flows		
% Citing matching	8.3	8.2
% Controls matching	6.2	5.9
Co-location premium	2.1	2.3
z-statistic	37.44	40.62
Co-location premium/% controls matching	0.34	0.39
n	414,511	410,372

Table 8. Spillover premiums (movers): prior co-location premiums for cross-field versus within-field flows (1990)

	Current location	Prior location
Cross-field flows		
% Citing matching	8.3	5.2
% Controls matching	4.6	3.2
Prior co-location premium	3.8	2.0
z-statistic	19.85	12.77
Prior co-location premium/% controls matching	0.82	0.63
n	33,618	33,618
Within-field flows		
% Citing matching	8.4	5.0
% Controls matching	6.4	3.6
Prior co-location premium	2.1	1.3
z-statistic	9.00	7.40
Prior co-location premium/% controls matching	0.33	0.36
n	26,116	26,116

co-location premium (proportion of citing matches less the proportion of control matches) is greater for the across-field flows. For example, Table 7 illustrates that the across-field co-location premium is 3.8 compared to only 2.1 for within-field.

It is perhaps more revealing to look at the ratio of the co-location premium to the percentage of matches in the control sample. As discussed in Section 2, using Bayes Rule, this ratio can be interpreted as the proportionate increase in the probability of a

however, we limit attention to citations that are in a different field from the originating patent, our control citations will also be (by construction) in a different field. Thus, we are less likely to see location matches for this subset of the data.

citation that is associated with co-location. Looking at Table 7 (1990 data), we see that co-location results in a 76% increase in the probability of a cross-field citation but only a 34% increase in the probability of a within-field citation.

Table 8 reports the results for movers. Prior co-location results in a 63% increase in the probability of a cross-field citation but only a 36% increase in the probability of a within-field citation. A similar pattern is found for current co-location (82% versus 33%). Thus co-location, both current and prior, does indeed appear to be most important in supporting knowledge flows when inventors are working in *different* fields.[35]

5. Conclusion

Much has been written about the 'death of distance.' Modern information and communications technologies are thought to have diminished the obstacles to economic interaction created by geographic separation. Yet the tendency for high technology industries (where knowledge-intensive outputs are essentially weightless) to be geographically clustered[36] suggests that proximity to sources of knowledge flows as inputs to R&D is critically important.

In this paper, we posit that geographic proximity works to overcome social distance and, once relationships are established, individuals can remain socially close even when they become geographically separated. We explore empirically how the prospect of separation affects the extent and form of social relationships that individuals develop and find evidence to support the hypothesis of an enduring social relationships effect with respect to knowledge flows.

We think these results are interesting in the context of three literatures. First, they provide additional insight into the processes through which economic knowledge diffuses. The results are consistent with the conjecture that social relationships facilitate knowledge spillovers, and they show how the geographic distribution of social capital may impact economic growth patterns in sometimes subtle ways.

Second, the results are relevant in the context of measuring what an economic location loses when a portion of its skilled workforce leaves. The migration literature has suggested the importance of knowledge spillovers to the gains and losses of locations from mobile labor (e.g. Borjas, 1995). But these effects generally have been viewed as unmeasurable. The JTH findings show that knowledge spillovers are geographically localized, which suggests an important source of location-specific loss when inventors leave. Our results suggest, however, that the losing location can nonetheless retain some degree of favored access to the knowledge generated by the departed inventor from their new location.[37]

35 We hasten to add that this does not imply that co-location leads to a larger absolute increase in the probability of a citation for cross-field inventor–patent pairs. The reason is that the unconditional probability of a citation is likely to be smaller for cross-field pairs than for within-field pairs (see Equation (1) above).

36 See, for example, the evidence in Audretch and Feldman (1996).

37 Interest in estimating the losses from the out-migration of skilled workers (a.k.a. the 'brain drain') has been growing as the competition for talent between and within national (or regional) economies has increased (Kapur and McHale, 2005).

Third, our results may be of interest to those studying the links between labor mobility and social capital accumulation (e.g. Glaeser et al., 2002). Using patent citation data as a proxy for knowledge flows and modeling knowledge flows as being facilitated by social relationships, our work shows how the rich data that is available on the geographic locations of patenting and citing inventors can be used to empirically examine how prospective mobility affects social capital accumulation.

Overall, we think these results are interesting in the context of increasingly knowledge-based and geographically mobile societies. But we have only touched on the interesting questions they raise. What does an economic region lose when a portion of its highly skilled workforce leaves—but is not (completely) forgotten? What does an economic region gain when it attracts highly skilled workers who remain networked to their former peers? How does the shift to a more mobile society affect individual incentives to develop economically useful social relationships? What are the implications of increased intranational and international mobility for the diffusion of technological knowledge and thus for regional and national government incentives to fund research and development? Answers to these questions may have great significance for policymakers interested in regional differences in growth and prosperity as well as for individuals or firms making privately optimizing location decisions.

Acknowledgements

We gratefully acknowledge thoughtful insights on working drafts of this paper from Erik Brynjolfsson, Wes Cohen, Bill Cooper, Robert Lowe, Bhaven Sampat, Scott Stern, Marie Thursby, Arvids Ziedonis, participants of the NBER productivity seminar and the Roundtable on Engineering and Entrepreneurship Research at Georgia Tech, and two anonymous referees. We also thank Alex Oettl and Raghav Misra for excellent research assistance. All errors and omissions are our own. This research was funded by the Social Sciences and Humanities Research Council of Canada (Grant No. 501-2001-0029). Their support is gratefully acknowledged.

References

Agrawal, A., Cockburn, I., McHale, J. (2003) Gone but not forgotten: labor flows, knowledge spillovers, and enduring social capital. Working Paper 9950, National Bureau of Economic Research.

Agrawal, A. (2006) Engaging the inventor: exploring licensing strategies for university inventions and the role of latent knowledge. *Strategic Management Journal*, 27: 63–79.

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

Audrestch, D., Feldman, M. (1996) R&D flows and the geography of innovation and production. *American Economic Review*, 86: 630–640.

Borjas, G. (1995) The economic benefits of immigration. *Journal of Economic Perspectives*, 9: 3–22.

Brown, J., Duguid, P. (1991) Organizational learning and communities of practice: towards a unified view of working, learning, and innovation. *Organization Science*, 2: 40–57.

Burt, R. S. (1992) *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Cockburn, I., Kortum, S., Stern, S. (2002) Are all patent examiners equal? the impact of examiner characteristics on patent statistics and litigation outcomes. Working Paper 8980, National Bureau of Economic Research.

Crane, D. (1965) Scientists in major and minor universities: a study of productivity and recognition. *American Sociological Review*, 30: 699–714.

Crane, D. (1969) Social structure in a group of scientists: a test of the "invisible college" hypothesis. *American Sociological Review*, 34: 335–352.

Dasgupta, P., David, P. (1987) Information disclosure and the economics of science and technology. In G. Feiwel (ed.) *Arrow and the Ascent of Modern Economic Theory*. New York, NY: New York University Press.

Dasgupta, P., David, P. (1994) Toward a new economics of science. *Research Policy*, 23: 487–521.

Glaeser, E., Laibson, D., Sacerdote, B. (2002) An economic approach to social capital. *The Economic Journal*, 112: 437–458.

Granovetter, M. (1973) The strength of weak ties. *American Journal of Sociology*. LXXIII: 1360–1380.

Jaffe, A., Trajtenberg, M. (2002) *Patents, Citations, and Innovations*. Cambridge, MA: The MIT Press.

Jaffe, A., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge flows as evidenced by patent citations. *Quarterly Journal of Economics*, CVIII: 577–598.

Jaffe, A., Trajtenberg, M., Fogarty, M. (2002) The meaning of patent citations: report on the NBER/Case Western Reserve survey of patentees. In A. Jaffe M. Trajtenberg (eds.) *Patents, Citations, and Innovations*. Cambridge, MA: The MIT Press.

Kapur, D., McHale, J. (2005) *Give us your best and brightest: the global hunt for talent and its impact on the developing world.* Washington DC: Brookings Institution Press for the Center for Global Development.

Krugman, P. (1991) *Geography and Trade*. Cambridge, MA: The MIT Press.

Lave, J., Wenger, E. (1993) *Situated Learning: Legitimate Peripheral Participation*. New York, NY: Cambridge University Press.

Marshall, A. (1890) *Principles of Economics*. London, UK: Macmillan.

Porter, M. (2000) Location, competition and economic development: local clusters in a global economy. *Economic Development Quarterly*, 14: 15–34.

Romer, P. (1986) Increasing returns and long-run growth. *Journal of Political Economy*, XCIV: 1001–1037.

Romer, P. (1990) Endogenous technological change. *Journal of Political Economy*, XCVIII: 71–102.

Singh, J. (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51: 756–770.

Song, J., Almeida, P., Wu, G. (2003) Learning-by-hiring: when is mobility more likely to facilitate interfirm knowledge transfer? *Management Science*, 49: 351–365.

Thompson, P., Fox-Kean, M. (2005) Patent citations and the geography of knowledge spillovers: a reassessment. *American Economic Review*, 95: 450–460

Trajtenberg, M., Jaffe, A., Henderson, R. (1997) University versus corporate patents: a window on the basicness of invention. *Economics of Innovation and New Technology*, 5, 19–50.

Trajtenberg, M. (2004) The names game: using inventors patent data in economic research. Mimeo: Tel Aviv University.