Foreignness and the Diffusion of Ideas

David M. Waguespack
SUNY Buffalo

Jóhanna Kristín Birnir
SUNY Buffalo

Introduction

Scholars of innovation frequently consider the diffusion of an idea, via metrics such as citations to scientific articles or citations to patents, as a post-hoc measure of the quality or economic importance of that idea. Thus good ideas are seen to eventually become the foundation for much follow-on work. At the same time it is well understood that ideas do not flow seamlessly across geographic space. In studies in which the United States is the focal country there are strong localization/clustering effects, such that geographic distance forms a barrier to the efficient assimilation of new information (Jaffe, Trajtenberg, & Henderson, 1993; Thompson & Fox-Kean, Forthcoming). By implication, new ideas that originate in geographically isolated regions will therefore diffuse less widely and/or rapidly. Alternatively, scholars of technology have theorized and found evidence that scientific and technological breakthroughs are most likely to occur when novel combinations of input are combined (Fleming, 2001; March, 1991). In complementary fashion, there is extensive work in social network theory indicating that while much of the above localization effects are explained by inter-personal social networks, the most valuable network positions are those that connect otherwise disconnected clusters (Burt, 2001; Granovetter, 1973; Powell, Koput, & Smith-Doerr, 1996). From this perspective, new ideas that are collaboratively developed by individuals in otherwise disconnected regional clusters are therefore more likely to result in truly novel ideas and therefore will diffuse more widely and/or rapidly.

The notions that localization effects will impair knowledge flow, while novel combinations result in higher quality and hence enhance diffusion, are in no apparent conflict. Assessing these effects simultaneously, however, requires a research design that

explicitly considers geographically remote collaborations, and distinguishes between the effects of simple geographic distance and the effects of crossing political/institutional borders on knowledge flow. On the collaboration issue, for instance, the studies of localization effects cited above assign all inventions to a single geographic locale and discard information on remote collaborations. On the issue of distance and political borders, it is typical for studies using US patent data to note in passing that approximately one half of modern US patents are of foreign origin, and that there is a liability to foreignness in the expected citation rate. However, despite the fact that a near majority of US patents are of non-US origin, no single nation comes close to the overall absolute contribution from the US, and therefore it is not surprising that in aggregate foreign invented US patents have lower citation rates, even from other foreign inventors, because each individual nation is in fact geographically isolated from the dominant region of US patent production and from each other. Although there are a series of more recent studies examining knowledge diffusion in multinational corporations with internationally dispersed research and development attempting to tap clusters of expertise (Almeida, 1996; Frost, 2001; Nobel & Birkinshaw, 1998; Singh, 2004; Zhao, 2003), what is not known is whether any residual foreignness liability for knowledge diffusion is a constant or varies with legal and political institutions.

We develop and test two sets of hypotheses related to geographically dispersed inventors and the diffusion of ideas using US patent data. First, and consistent with both extant research on geographic concentration and weak ties, we find that US patents with dispersed US inventors garner higher rates of citation. Furthermore, patents involving US/non-US inventor collaborations diffuse at the same rate as those developed in a single

location in the US, indicating that the positive effect of remote collaboration is offset by the difficulty of integrating ideas partially developed in a different legal and institutional context (relative to the US, which is the focal setting when using US patent data). Isolated foreign patents, finally, diffuse at the lowest rates.

In the second set of hypotheses we focus exclusively on US patents of foreign origin, examining the rate at which these ideas diffuse across international borders to US inventors as a function of the legal system and socio-economic status of the country of origin. Legal precepts related to intellectual property vary dramatically across nations even on such seemingly basic concepts as priority and novelty, and US patents only directly confer intellectual property rights within the United States. We find that rates of diffusion to the US vary with legal family, and are greatest for inventions originating in those nations most congruous with US practices.

Invention Diffusion and Foreignness

Citation patterns among US patents are commonly used to measure knowledge flows (Jaffe et al., 1993; Sorenson & Fleming, 2001; Thompson et al., Forthcoming). New inventions seeking patent protection are legally required to cite the technology upon which they build, and as such represent a way to establish linkages between distinct patents. Conversely, citing another patent as 'prior art' limits the scope of novelty on a new invention, so inventors have an incentive to minimize these citations (Sorenson et al., 2001). Thus, highly cited patents are described variously as having diffused widely, as representing more important inventions, or as serving as a platform for technology development.

One of the most widely known papers analyzing patent citation patterns finds that knowledge spillovers are geographically localized (Jaffe et al., 1993). The probability of knowledge exchange decreases without physical proximity, although whether this is a result of complex technical information being difficult to transmit without direct personal interaction (i.e. inventors living in the same are more likely to know one another personally) or because new ventures are more likely to start in areas where clusters of expertise already exist is a question of continuing debate (Sorenson & Audia, 2000). Another common finding, which is generally not discussed in great details in studies on knowledge flow and patent citations, is that foreignness, as in US patents invented overseas, imposes an additional liability on the probability of knowledge diffusion.

In itself the finding that knowledge flow tends to be localized is not surprising, as it resonates both with the idea that technical/scientific information has a tacit character, and with the notion that densely connected social networks will tend to form. There is a seeming conflict here, however, with the notion that novel combinations of inputs also result in the most novel inventions (March, 1991). Granoveter (Granovetter, 1973) noted that in densely connected clusters most information exchange is redundant, and hence the most valuable information comes from sources outside the cluster. In logical extensions, other have noted that boundary spanning results in more innovative output (Powell et al., 1996), that gatekeepers between organizations and external environments strongly influence an organization's engineering efforts (Allen, 1978), and that information brokers mediating between disconnected nodes reside in uniquely powerful positions (Burt, 1992; Burt, 2001).

Combined with the knowledge that patent citations cluster geographically, these insights suggest that collaborations among geographically dispersed inventors will result in more novel inventions. Table one explores this issue, examining citation rates for US patents based on inventors geographic dispersion and type of assignee. The table shows that in general citations rates are lowest for patents with only foreign inventors (*Non-US only*), and that both within-US collaboration (*2+ US States*) and US-international collaboration (*US and Non-US*) have higher rates of citation than US patents developed in a single US state (*Single US state*). The foreignness liability holds for all non-governmental assignees, but the premium associated with international collaborations disappears for non-US firms and US university assignees.

[Insert table 1 about here]

Building on the recognition that foreign clusters of expertise represent an opportunity for multi-national firms engaged in research and development efforts, a series of recent papers have examined this phenomenon, with a substantive focus mostly on how organizational structures can mitigate concerns with misappropriation of overseas research and development efforts (Almeida, 1996; Frost, 2001; Nobel et al., 1998; Singh, 2004; Zhao, 2003). For their part, and as is common with patent related research, Jaffe et al. subsume all patents with geographically dispersed origin into single focal location based on the plural address. Thus in most cases, and despite the knowledge the invention location matters to diffusion, geographic collaboration patterns are neglected. These observations lead us to propose the following hypotheses:

H1: patents with geographically isolated inventors diffuse less widely

H2: patents with geographically dispersed inventors diffuse more widely

Political Borders and the Flow of Knowledge

The question left unanswered above is whether there is variation in how widely patents with foreign inventors diffuse to US inventors. Are there systematic differences in how likely an US patent is to be cited by other patents filed in the US depending on where the invention in the patent to be cited originates? Others have argued that, cross-nationally, institutions affect economic development of which intellectual property rights are a sub-category. For instance, La Porta et. al. (1999) argue that"legal families" diverge in their effects on economic development. More specifically, they posit, that "poor" economic performance is the result of "poor" governance which in turn is associated with, among other things, French civil laws. The reason, they explain, is that common law developed as a defense for Parliament, individuals, and property owners against an encroaching sovereign, whereas the civil law developed more as an instrument of the sovereign. La Porta et. al. (1998) further support this view by articulating a specific example of greater investors rights in common law systems with consequent positive results for the national economies. Birnir, Waguespack and Shroeder (2004), in turn, apply the idea that local institutions matter for rates of foreign patents sought in the United States.

The difference between La Porta's et. al. (1999 and 1998) argument and ours is that we do not look at national patenting rates, rather we examine only diffusion of US

patents with foreign inventors to US based inventors. Our argument also differs from Birnir, Waguespack and Schroeder in that we do not examine the longevity of institutions and their effect on general policy stability with implications for patenting propensity but the type of institution and their effect of the diffusion of ideas. We argue that the assimilation US inventors seek via the construction of ties to inventions that originate remotely depends, in part, on the remote institutional context even though the patent for the foreign invention is filed in the US. The reason, we posit, is that two determinants of a US inventor's willingness to cite a remotely developed patent are her understanding of the information contained in the patent she cites and the increase of risk in her legal exposure that results from citing the remote invention. Legal institutions and their effect on scientific development differ significantly between countries**.** This in turn influences the information contained in any particular patent even when the patent is sought in the United States. Therefore, even though the cites we discuss in this paper are all of patents granted in the US, a US inventor's understanding of the information that is conveyed by a remotely developed patent decreases as the institutions under which the remote patent was originally developed differs from the common law system that governs patent law in the US.

The increased legal exposure that an inventor incurs by citing patents of inventions that were developed remotely as opposed to patents developed locally is equally significant. One of the most important, if not the most important principle of patent law internationally is the principle of territoriality. According to the principle of territoriality an inventor’s patent rights are only protected in the country where the patent was granted (Mills III et al. 2002). Therefore, an inventor who applies for and is granted

a patent in any one country *does not* have patent protection in any other country. International conventions such as the Paris Convention of 1833 and the 1978 Patent Cooperation Treaty oblige signatories to grant applicants domiciled in whatever state of the union whatever its own patent law accords to its own nationals and allows a patentee to file one application and have it become a national application in as many countries as she designates respectively. It is, however, up to the individual country to evaluate and accept or reject the application.

Consequently, a US inventor that applies for a patent in the United States does not have any patent rights elsewhere. If that inventor establishes remote ties to foreign inventors by citing remotely developed inventions that were patented in the United States she increases the remote exposure of her US patent. If this increased exposure occurs in a country where patenting institutions are more liberal than patenting institutions in the US in that it is easy to obtain a patent, the risk in her exposure increases because someone else could easily obtain a patent of her invention in that remote location. Furthermore, the more dissimilar the legal channels under which patenting rights are pursued the less likely that she has any legal recourse and the greater the risk of her exposure. Along the lines of la Porta et. al. (1999) we, therefore, posit that legal families matter to inventors incentives to establish ties through citations of patents of inventions developed remotely. Table 2, reporting citation rates from US inventors for US patents of foreign origin, seems to provide a priori support for the idea that legal family matters. To more rigorously assess this relationship, we propose our third hypothesis that:

H3: Legal families affect an inventors incentive to cite patents based on inventions

developed remotely.

[Insert table 2 about here]

However, we seek to move beyond the idea that the reason legal families affect economic activity is the general notion of a system development of the legal family in response to a sovereign. Rather we make a more specific argument as we posit that incentives to build ties across and within legal families differ because of specific institutions that influence the information contained in the remotely developed patent that is cited and the change in the risk of exposure that the inventor incurs by citing a remotely developed patent. The legal family distinction made by La Porta et. al. (1999) is between Civil and Common law. We begin by explaining why we believe that civil and common law systems affect understanding and risk in exposure differently. Then we explain why and how the effect of other legal families also differs.

According to Rosenberg (1975) seven principal features distinguish between the essential characters of civil and common law patenting systems around the world. The extent to which applications are screened prior to grant of patent rights has the most significant effect on the character of the system. The element that distinguishes between screenings of applications prior to the granting of a patent is the novelty requirement. Countries that adhere to the English common law tradition such as the United States commonly pass a judgment on the novelty of an invention and are said to have an examination system. To the contrary, countries that followed the Roman civil law tradition such as France and Germany, do not make a novelty determination and are said

to have a registration system. In 1995 the WTO replaced the General Agreement on Trade and Tariffs (GATT) as the organization overseeing the multilateral trading system. Of these agreements the TRIPs (Trade Related Aspects of Intellectual Property Rights) provides detailed rules for minimum levels of protection for intellectual property in WTO member countries. The creation of the WTO standardized some basic rules regarding novelty and prior art for both legal systems, yet legal traditions are not fully merged as result of the agreement. Therefore, the effect of divergent legal families clearly differed prior to 1995 and there is good reason to believe the effect still differs to some extent.

How then does the difference between common law examination requirements and civil law registration requirements affect the incentives of a US inventor to cite inventions developed remotely but patented in the US? The US is a common law system that adheres to the examination requirement. Therefore, it is common knowledge that any invention developed under this system must meet the more stringent examination requirement before receiving a patent. This likely affects the development of the invention and documentation at all stages of the inventive process. A US inventor, therefore, likely has a better understanding of the process under which inventions in other common law systems are developed. More importantly, however, since the examination requirement is more stringent than the registration requirement, patents are likely more easily obtained in registration systems. Consequently, a US inventor's risk in exposure by establishing ties is greater in civil law than common law systems:

H4. Inventors who develop a patent in the US common law system are less likely to cite patents developed in civil law systems than in other common law systems.

Common law and civil law are but two of the many legal families in the world. While different, common law and civil law systems share certain fundamental principles that distinguish them significantly from other legal families. This includes the emphasis on the individual right to patent. While Socialist legal families in some cases incorporate elements of civil and/or common law patent legislation others do not (Reynolds 2004). Consequently, we posit that on both accounts, understanding and risk in exposure, inventors filing a patent in the US have an even lower incentive to cite patents developed under socialist systems. Therefore:

H5: Inventors who develop a patent in the US common law system are less likely to cite patents developed in socialist systems than either common law or civil law systems.

Data and Methods

We test our hypotheses using data on US patents granted from 1990 to 1994 (Hall, Jaffe, & Trajtenberg, 2001). To this data we added adjusted assignee identifier codes, distinguishing US Firms from US Universities, based on a list of patenting Universities published by the US Patent and Trademark Office (USPTO 2002). For patents with foreign inventors we supplemented the patent data with attributes of the inventor's nation(s). These additional attributes include coding of language and legal families developed by the authors, selected socio/economic indicators from the World Bank (World Bank 2002), and the index of local intellectual property rights developed by

Ginarte and Park (Ginarte & Park, 1997).

Tables 3 and 4 describe the analysis data in greater detail. We take essentially two different slices from the patent data, one for all US patents issued between 1990-1994, and another that further restricts this sample to those of (partially or completely) non-US origin. Starting with Table 3 for the larger set of all 1990-1994 patents, the dependent variable *Cites Received* is a count of the number of future patent citations received for the 1990 to 1994 patent population. The variables of interest presented in Table 3 concern the geographic dispersion of inventors. *Non-US First Inventor* is a dummy variable indicating that the first inventor on the patent is not a US resident. The variables *US only single state*, *US only 2+ States*, *US and Non-US*, and *Non-US only* represent mutually exclusive categories for inventor collaboration pattern recoded as dummy variables, with *US only single state* serving as the reference category in estimation. As these variable labels suggest what is measured is whether the inventors reside in the same state, reside in different states, engage in an international collaboration, or reside outside the US.

[Insert table 3 about here]

Turning to Table 4, we report descriptive statistics for those cases where the US patent originates partially or completely outside the US. The dependent variable *Citations Received from US Inventors* counts only those future citations coming from patents with a US inventor in order to gauge when ideas developed abroad will diffuse back to the US. The variables of interest focus on direct inventor linkages to the US, and

the legal and socio/economic characteristics of the foreign inventor's nations. The variable *US States(n)* indicates how many US states are represented among the inventors on the patent. Recall that we only include patents with some foreign inventors and measure citations by patents with some US inventors, so our a priori assumption is that US/non-US collaboration will drastically enhance diffusion to the US. The variable *English Language* is a dummy indicating that the local native language is english. The variables arrayed under the *Legal System Origin* (*English, French, German, Scandinavian, Socialist, Roman-Dutch,* and *Islamic*) are dummy variables indicating the local legal code. These variables are non-mutually exclusive in that some international collaborations do occur, although these are rare and so we use *German* as the excluded reference category for legal family to avoid collinearity problems. The social and economic variables *Development (ln GPD/cap.), Education (% Tert. Enroll.), Scientific Publications,* and *Foreign Direct Investment* are all derived from the 2002 World Bank Development Indicators (World Bank 2002). Although interesting, we treat these indicators more as control variable because they are frequently missing, especially for developing nations and/or nations with disordered political situations. The variable *Ginarte/Park IPR Index* (abbreviated as *GP IPR Index* in later tables) comes from the eponymous index of intellectual property rights for 1960 to 1990 (Ginarte et al., 1997). This indicator allows us to evaluate the enforcement of intellectual property rights independently of congruity between the US and local legal system. Unfortunately this variable is also frequently missing for countries represented in the US patent data, particularly those that were under Communist control at the time of Ginarte and Park's writing. We use Ginarte and Park's 1990 IPR index score, and thus *GP IPR Index* is

missing for countries frequently mentioned in discussions of global innovation, such as Romania, Hungary, and Taiwan.

[Insert table 4 about here]

Finally, we have a series of control variables included in all specifications. *Claims* reports the number of distinct novel elements that the inventor(s) list as part of the invention. *Citations Made* is the number of citations made by the patent to other patents. Both *Claims* and *Citations Made* are included to control for the scope of the patent (i.e. the likelihood that it is relevant to any future inventor) and the extent to which the inventor(s) is already linked to other inventors. The *Assignee Type* array of dummy variables indicates the type of organization that owns the patent. Finally, all equations include sets of dummy variables for *Grant Year* (5 variables, 1990 to 1994), and US Patent and Trademark Office *Technology Classification* (411 classes for the full set of patents, 405 classes for the foreign origin set).

The dependent variables, two variants of future citations to the focal patent, are non-negative event counts and as such inappropriate for ordinary least squares estimation. The dependent variables also do not fit a Poisson distribution in that means and variances are not equivalent, so we instead estimate coefficients using negative binomial estimation (Hausman, Hall, & Griliches, 1984).

Results

The regressions in Table 5 test the first hypotheses relating to inventor isolation

and remote collaboration on invention diffusion. The case selection consists of all US patents that were issued from 1990 to 1994, with the independent variable measuring the number of future citations. Control variables for *Claims*, *Citations Made*, four *Grant Year Dummies* (with 1990 the reference year), 411 *Technology Class Dummies*, and five *Assignee Types* (with unassigned the reference category), are included in all specifications. With the exception on *Non-US Government Assignee Type*, these control variables are statistically significant in all specifications. The first model accounts for inventor dispersion with a simple coding of inventive location, *Non-US First Inventor*, based solely on the location on the first listed inventor. Based on this simple coding, which reflects the typical method for geographically partitioning US patents, inventions developed outside the dominant region of US patent production suffer significantly lower rates of diffusion. The second model examines the effects of collaboration across regions explicitly, breaking patents into one of four geographic collaboration patterns. Relative to the reference category *US only single State*, collaborations within the United States (*US only 2+ states*) diffuse most widely, there is no marginal difference for US/international collaborations (*US and non-US*), and a significant negative effect on the overall diffusion of inventions developed solely outside the US (*Non-US only*).

[Insert table 5 about here]

Specification 1 in Table five confirms the conventional wisdom regarding the effect of geographic barriers. We show that developing a patent remotely (*Non-US First Inventor*) is a liability for the diffusion of that idea within the United States even when

the patent is sought in the US. Furthermore, and as expected, most of the control variables are significant and have the expected sign. The number of *Claims* is highly significantly and positively associated with the rate of citation the patent receives, as is number of *Citations* made, and all of the *Grant Year* and *Technology Class* dummies. If the patent assignee is a *US firm*, a *Non-US Firm*, or a *US University* the rates of cites of that patent are highly significantly greater than rates of cites of patents that are unassigned or assigned to and individual. To the contrary when the assignee is the *US government* the rates of cites of that patent are highly significantly lower than rates of cites of patents that are unassigned or assigned to and individuals. Finally, the rates of cites of patents that are assigned to *Non-US Governments* is not statistically significantly different from rates of cites of patents that are unassigned or assigned to and individual**.**

The second model teases out the hypothesized beneficial effects of dispersion and juxtaposes these with the drawbacks of geographic barriers for dissemination of ideas. This model shows an increasing rather than absolute effect of geographic barriers that is counteracted by the benefits of dispersion. Only when geographic barriers include international borders do the drawback of remoteness outweigh the positive effects of collaboration. The variable accounting for dispersion of inventors over two or more states (*US only, 2+ States*) is positively and highly significantly associated with rates of cites of the patents when compared to rates of cites of patents developed in a single US state. Clearly, therefore, dispersion has benefits that outweigh geographic barriers within the US. Furthermore, the rates of cites of patents resulting from US and foreign collaborations *(US and Non-US*) are not significantly different from rates of cites of patents that are developed within a single US state. Here the geographic barriers are

greater but still do not outweigh the beneficial effects of dispersion completely. Finally, patents that originate remotely without any US collaboration (*Non-US only*) are significantly less likely to be cited than patents developed within a single US state. Here the geographic barriers clearly outweigh the potential benefits of citing a remote and therefore potentially more novel invention. All of the control variables retain their sign and significance in the second model when compared to the first.

These results show that dispersion of inventors has both positive and negative effects on the rates of cites of US patents. Some dispersion of inventors is good, presumably because of the positive effects of novelty that result from increased diversity in dispersed inputs and increases the value of the invention. Significant dispersion of inventors across national borders is subject to greater geographic barriers in dissemination of that information and where inventors are located exclusively outside the US the dissemination of their idea within the US is severely restricted. What these results do not show is whether there is a difference between the effects of different types of dispersion of inventors. Are US inventors more or less likely to cite a collaborative or a foreign invention depending on where the foreign inventor resides? According to hypotheses three, four, and five, this is indeed the case. We hypothesized that US inventors incentives to cite a patent that originates in part or exclusively in a remote location depends on the institutional system of that location. More specifically, we suggested that US inventors are most likely to cite patents that originate in other common law systems, less likely to cite patents that originate in part or wholly in civil law systems and least likely to cite patents that originate in socialist systems. The reason is that as the institutional influences on the invention diverge from the common law system the US

inventor has less understanding of the information contained in that patent. Furthermore, the US inventor increases her risk in exposure in both civil law and socialist systems when compared to common law systems due to differences in patent law. The regressions in table six test these hypotheses.

The cases included in the test in table six are US patents of foreign origin, with the dependent variable counting only the future citations received from US inventors. The independent variable of interest in the first model is a dummy variable accounting for countries where *English* is either the official language or the first language spoken by a majority of the population. The control variables in this model include the number of distinct novel *Claims* that the inventor(s) list as part of the invention, the number of *Citations* made by the patent to other patents, set of dummy variables for accounting for *Grant Year* of the patent, set of dummy variables accounting for *Technology Classification*, and *Type of Assignee* array of dummy variables that indicate the type of organization that owns the patent. The second model includes all the same control variables but drops the English language dummy in lieu of the seven non-mutually exclusive dummy variables accounting for the *Legal System Origin* of the patent (*English, French, German, Scandinavian, Socialist, Roman-Dutch*, and *Islamic*). The third model includes all of the variables in model two and adds a series of socio-economic controls including level of *Development* measured as GDP per capita, level of *Education* measured as percent of population enrolled in tertiary education, Number of *Scientific Publications*, *Foreign Direct Investment*, and the *GP IPR index* that measured foreign enforcement of intellectual property rights.

[Insert table 6 about here]

The control variables accounting for *Claims, Citations, Grant Year and Technology Class* are all highly statistically significantly and positively associated with rates of citations of patents in all three models. In reference to patents that are unassigned or assigned to an individual, rates of cites of patents assigned to *US Firms* and *US Universities* are highly statistically significantly greater. Rates of cites of patents assigned to the *US Government* are greater than are rates of cites of unassigned patents or patents assigned to an individual but this effect varies from being statistically insignificant to highly statistically significant between models. To the contrary, rates of cites of patents assigned to *Non-US Firms* are highly statistically significantly lower than rates of cites of unassigned patents or patents assigned to an individual in every model and the same is true for rates of cites for patents assigned to a *Non-US Government* but the significance of this effect disappears once we control for legal family.

The *English* language variable is highly statistically significant and positively associated with rates of citations of patents in the first model. In other words, US inventors are more likely to cite patents that originate in countries where English is primary language of communication. In the following model we attempt to discern whether this effect is simply due to facility in communication as a result of shared language, or whether variation in legal institutions makes a difference. To this end we introduce seven variables accounting for different legal families. We include six of these variables in the regression (*English, French, Scandinavian, Roman-Dutch, Socialist, Islamic*) in reference to the seventh which accounts for the *German* civil law variant. The

English common law variable is highly correlated with the English language variable and as expected US inventors are significantly more likely to cite a patent that originates in any of the other countries whose legal system is based on common law. If this effect were exclusively due to language we would not expect to see significant differences in the rates of citations of patents in non-english speaking countries (particularly not after we account for socio-economic factors in model three). What we find instead is that there are significant differences between rates of citations from different types of civil law systems and socialist systems. When we account for socioeconomic factors the *Roman-Dutch* common law system also achieves positive significance when compared to *German* civil law systems.

Clearly, the difference in rates of citations is related to the different legal families. Precisely what that relationship is, however, fairly complex. A few findings are robust whether or not we control for socio-economic indicators, and all differences between coefficients reported below are supported by likelihood-ratio tests. First, rates of cites of patents originating under *Islamic* law does not significantly differ from rates of cites of patents that originate under *German* civil law. This finding is likely due to the fact that there are very few systems where Islamic law is the sole foundation of the legal system of a country. Second, as expected, US inventors cite patents that originate in English common law systems more frequently than patents originating in any civil law system (*French, German, Scandinavian*). Third, patents originating in *Scandinavian* civil law systems get more citations from US inventors than those from *German* civil law systems and *French* civil law systems. The reason for this finding is likely that patent rights are on average high in Scandinavia, whereas German and French civil law systems include

countries in Africa, Latin America, Eastern Europe and Asia where patent rights tend to be lower. Therefore, we explicitly control for patent rights (*GP IPR Index*) this finding strengthens. Jointly these results support the second and third hypotheses that legal families affect citations and that inventors in common law systems are less likely to cite patents originating in civil law systems.

The socioeconomic controls show that development and legal enforcement of patent rights matter for diffusion of ideas. *Development* measured in GDP per capita is highly significant and positively related to rates of US cites of patents developed in foreign countries. *Education* is not significantly related to rates of cites, but the presence of highly educated researchers engaged in original scientific research (*Scientific Publications*) is positive and significant. Interestingly *Foreign Direct Investment* does not significantly affect the rates of cites of patents developed outside the US. The likely reason for this is that foreign direct investment might affect the number of ideas that emerge remotely but is less likely to affect the subsequent dissemination of that idea in the US. The quality of local property rights measured through the *GP IPR index* is highly significant and negatively related to citations of patents. This result is somewhat surprising, but seems to indicate legal precepts and actual practices are possibly distinct. Ultimately we are loathe to overinterpret the intellectual property rights index and other socio-economic variables as nearly 10% of cases are missing on these measures.

The socioeconomic control factors affect the remaining legal family categories. Before controlling for socio-economic indicators rates of cites of patents that originate in Socialist countries are also significantly lower than rates of cites in either common law or any of the civil law countries, supporting the third hypothesis that US investors are less

willing to risk exposure in socialist countries than in either common law or civil law countries. This finding disappears once we control for socioeconomic factors but the reason is most likely that the data accounting for quality of property rights does not include indicators for socialist or dependent countries and those are all dropped from the sample as indicated by the reduced number of cases in the third model. Therefore, model two accounts more accurately for the effect of *Socialist* legal systems than does model three. Furthermore, the reason the *Roman-Dutch* system becomes significant is that the relative significance of these countries (Botswana, South Africa and Sri Lanka) increases when other countries drop out indicators and the Roman-Dutch system is a variant on a common law system. (The fourth Roman-Dutch system, Namibia drops out when we control for socioeconomic.) This finding, in turn, further supports the idea that the difference between rates of cites depending on where the patent originates are not only due to language because Sinhala is the official language in Sri Lanka and while English is the official language in Botswana, it is only one of eleven official languages in South Africa.

Finally, rates of cites of patents that originate in *German* and *French* civil law systems differ but how they differ changes once we control for socio-economic factors. We believe that the most significant difference between *French* and *German* civil law systems between the two models is that in the second model *French* civil law systems are disproportionately located in Africa and Latin America. *German* civil law systems, however, tend to be in Eastern Europe in addition to the more developed areas of Asia in Japan, Korea and Taiwan. In the second model *French* civil law systems, therefore, seem to perform worse than *German* civil law systems. Once we control for quality of patent

rights in model three, *French* civil law systems seem to perform better than *German* civil law systems. The most likely explanation for this difference, is not, however, that *French* civil law institutions are better understood and considered to provide lower risk in exposure than *German*. Rather the most likely explanation is that the overwhelming majority of Latin American countries use a *French* civil law system and the geographic proximity and cultural and linguistic ties between the multitude of Latin American residents and immigrants in the United States provide additional social ties that encourage diffusion of ideas that originate in *French* civil law systems above and beyond that of *German* civil law systems.

In sum, therefore, differences in legal families in the country of origin of an invention patented in the US clearly affect diffusion of the ideas to US inventors as we proposed in hypothesis two. Furthermore, US inventors are overall less likely to cite patents that originate in civil law countries than in other common law countries as predicted in the third hypothesis. In addition, US inventors are less likely to cite patents originating in socialist countries than patents that originate in either common law or civil law countries as proposed in the fourth hypothesis. We did not, however, uncover any evidence that suggests there are institutional differences between divergent types of civil law systems above and beyond what can be explained by patent rights enforcement and geographic proximity.

Conclusion.

In this paper we attempted to reconcile the conventional wisdom that knowledge tends towards geographic clustering with the conventional wisdom that breakthrough

science and technology is more likely to result when otherwise disparate spheres of information are combined. In analysis of the citations received by US patents, a common measure of the technical and economic importance of technology, we find that cross-state collaborations do indeed result in higher rates of citation relative to new inventions where all inventors are located in the same region. Furthermore, the liability associated with foreignness, the lower rate of diffusion expected for inventions developed solely outside the United States, is overcome in cases of US/international collaboration. We take this result to indicate that, in the case of US/international collaboration, the net positive benefit to remote collaboration is offset by a constraint on assimilating technology developed in a different legal setting. Finally, in order to determine whether the constraint with crossing political borders is constant or variable, we examined this issue of diffusion across international borders in more detail, examining citations from US inventors to US patents with a non-US origin. As expected, we found that citation rates are highest for US patents from nations with legal systems most consonant with US institutions.

Before concluding it is important to acknowledge the limitations that we are aware of with the analysis presented here. First, US patent citations are somewhat noisy measures of knowledge flow in that citations can be placed in the document by the inventor, the patent attorney, or the patent examiner (Thompson, 2004). Obviously it is only when the inventor makes a citation that an input to the creative process can be said to have occurred, but assuming these non-inventive citations are distributed independently of the effects we articulate we see no consequent bias in our results. Secondly, an alternative explanation of the finding that dispersed collaboration results in

greater citations is that distinct local diffusion effects were simply multiplied. While this is plausible, we also allow sufficient time for future citations to occur, between 5 and 10 years depending on the particular patent grant year, that we expect truly valuable ideas to have sufficient time to expand beyond the local setting. Thirdly, an alternative interpretation of the lower returns to patents developed under legal systems different from the US standard may be that there is a first order effect in that under-investment and hence intrinsically lower quality invention occurs in those places with less compatible legal systems. If so we expect that controlling for assignee type, foreign direct investment, and local development makes our results robust. Each of these concerns merit further consideration, and as such we also present them as promising directions for future research.

References

Allen, T. J. 1978. ***Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organizations***. Cambridge: MIT Press.

Almeida, P. 1996. Knowledge sourcing by foreign multinationals: Patent citation analysis in the US semiconductor industry. ***Strategic Management Journal***, 17: 155-165.

Birnir, J. K., D. Waguespack, & J. Shroeder. (2004) Regimes, Veto Players and Technological Development: Patenting in Latin America and the Caribbean. ***Working paper***.

Burt, R. S. 1992. ***Structural Holes***. Cambridge, MA.: Harvard University Press.

Burt, R. S. 2001. Structural Holes versus Network Closure as Social Capital. In N. Lin, K. S. Cook, & R. S. Burt (Eds.), ***Social Capital: Theory and Research***: 31-56. New York: Aldine de Gruyter.

Fleming, L. 2001. Recombinant uncertainty in technological search. ***Management Science***, 47(1): 117-132.

Frost, T. S. 2001. The geographic sources of foreign subsidiaries' innovations. ***Strategic Management Journal***, 22(2): 101-123.

Granovetter, M. S. 1973. Strength of Weak Ties. ***American Journal of Sociology***, 78(6): 1360-1380.

Hausman, J., Hall, B. H., & Griliches, Z. 1984. Econometric-Models for Count Data with an Application to the Patents R and D Relationship. ***Econometrica***, 52(4): 909-938.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. ***Quarterly Journal of Economics***, 108(3): 577-598.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer & Robert W. Vishny. 1998. "Law and Finance". The ***Journal of Political Economy***, 106(6):1113-1155.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer & R. W. Vishny. 1999. "The Quality of Government." ***The Journal of Law, Economics, and Organization,*** 15(1): 222-279

March, J. G. 1991. Exploration and Exploitation in Organizational Learning. ***Organization Science***, 2: 71-87.

Mills III, John G. et. al. 2002. ***Patent Law Fundamentals. Second Edition.*** Thomson West. New York.

Nobel, R., & Birkinshaw, J. 1998. Innovation in multinational corporations: Control and communication patterns in international R & D operations. ***Strategic Management Journal***, 19(5): 479-496.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. ***Administrative Science Quarterly***, 41(1): 116-145.

Reynolds, T. H. & A. A. Flores. 2004. ***Current Sources of Codes and Basic Legislation in Jurisdictions of the World. Foreign Law Guide.*** University of California, Berkeley. Berkeley, California. At:www.foreignlawguide.com

Rosenberg, P. D. 1975. ***Patent Law Fundamentals***. Clark Boardman Company. New York.

Singh, J. 2004. Multinational Firms and Knowledge Diffusion: Evidence using Patent Citation Data. ***Working paper***.

Sorenson, O., & Audia, P. G. 2000. The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940-1989. ***American Journal of Sociology***, 106(2): 424-461.

Sorenson, O., & Fleming, L. 2001. Science and the Diffusion of Knowledge. ***Working paper***.

Thompson, P. 2004. Patent Citations and the Geography of Knowledge Spillovers: What do Patent Examiners Know? ***Working paper***.

Thompson, P., & Fox-Kean, M. Forthcoming. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. ***American Economic Review***.

World Bank. 2002. World Bank Development Indicators CD-Rom.

US Patent And Trademark Office. 2002. US Colleges And Universities-Utility Patent Grants, Calendar Years 1969-2000. ***http://www.uspto.gov/web/offices/ac/ido/oeip/taf/univ/univ_toc.htm***.

Zhao, M. 2003. Doing R&D in Countries with Weak IPR Protection:Can Corporate Management Substitute for Legal Institutions? ***Working paper***.

Table 1: Mean Citation Rate by Assignee Type and Collaboration Pattern for US Patents 1990-1994.

Assignee Type	Geographic Collaboration Pattern				
(n) (mean)	Single US state	2+ US States	US and Non-US	Non-US only	Total
Individual	60,763 4.55	2,386 6.20	520 5.93	22,863 3.60	86,532 4.35
US Firm	155,972 6.55	21,701 7.25	4,238 7.11	12,108 5.08	194,019 6.55
Non-US Firm	1,967 5.27	272 6.87	1,740 5.10	184,577 4.52	188,556 4.54
US Government	4,698 3.69	926 3.99	113 4.96	15 6.27	5,752 3.77
Non-US Government	5 3.80	1 1.00	18 6.67	2,064 3.99	2,088 4.01
US University	5,724 7.12	1,158 7.08	458 6.85	21 5.62	7,361 7.10
Total	229,129 5.96	26,444 7.03	7,087 6.48	221,648 4.45	484,308 5.34

Table 2: Mean Citation from US Inventors by Legal System for US Patents 1990-1994 with Non-US Invention Origin.

Legal Family[a]	Citations from US Inventors	
	n	Mean
German	161,847	2.08
English	29,672	3.14
French	30,923	2.13
Scandinavian	6,925	2.44
Socialist	742	2.18
Roman	547	2.40
Islamic	258	2.99

[a] For this table patents with inventors in multiple legal families are reported once in each category.

Table 3: Descriptive Statistics, All US Patents 1990-1994.

	n = 484308	
	Mean	St. Dev.
DEPENDENT VARIABLES		
Citations Received	5.34	7.71
INDEPENDENT VARIABLES		
Claims	12.78	10.51
Citations Made	8.06	7.70
Assignee Type:		
Individual Assignee	0.18	0.38
US Firm	0.40	0.49
Non-US Firm	0.39	0.49
US Government	0.01	0.11
Non-US Government	0.00	0.07
US University	0.02	0.12
Non-US First Inventor	0.46	0.50
Inventor Dispersion Pattern:		
US only single state	0.47	0.50
US only 2+ States	0.05	0.23
US and Non-US	0.01	0.12
Non-US only	0.46	0.50

Table 4: Descriptive Statistics, All US Patents 1990-1994 with Non-US Invention Origin.

	n = 228732	
	Mean	St. Dev.
DEPENDENT VARIABLES		
Citations Received from US Inventors	2.23	4.10
INDEPENDENT VARIABLES		
Claims	11.10	9.19
Citations Made	6.06	5.05
Assignee Type:		
Individual Assignee	0.10	0.30
US Firm	0.07	0.26
Non-US Firm	0.81	0.39
US Government	0.00	0.02
Non-US Government	0.01	0.09
US University	0.00	0.05
US States (n)	0.04	0.21
English Language	0.12	0.33
Legal System Origin:		
German	0.71	0.45
English	0.13	0.34
French	0.14	0.34
Scandinavian	0.03	0.17
Socialist	0.00	0.06
Roman	0.00	0.05
Islamic	0.00	0.03
Socio/Economic Indicators:		
Development (ln GPD/cap.)[a]	24.15	0.44
Education (% Tert. Enroll.) [a]	39.38	13.65
Scientific Publications [a]	30154.14	12308.59
Foreign Direct Investment [a]	0.52	0.86
Ginarte/Park IPR Index [b]	3.79	0.35

[a] n = 219176 for table 6.
[b] n = 204446 for table 6.

Table 5: Negative Binomial Estimation
Dependent Variable = Citations Received
Case Selection: All US Patents Issued 1990-1994

	(1)	(2)
Claims	0.006	0.006
	(0.000)**	(0.000)**
Citations Made	0.010	0.010
	(0.000)**	(0.000)**
Grant Year Dummies	(4, **)	(4, **)
Technology Class Dummies	(411, **)	(411, **)
Assignee Type[a]*:*		
US Firm	0.140	0.138
	(0.004)**	(0.004)**
Non-US Firm	0.109	0.104
	(0.006)**	(0.006)**
US Government	-0.149	-0.154
	(0.014)**	(0.014)**
Non-US Government	-0.013	-0.017
	(0.022)	(0.022)
US University	0.212	0.208
	(0.011)**	(0.011)**
Non-US First Inventor	-0.151	
	(0.005)**	
Inventor Dispersion Pattern[b]*:*		
US only, 2+ States (1 = yes)		0.049
		(0.006)**
US and Non-US		0.006
		(0.011)
Non-US only		-0.144
		(0.005)**
Constant	0.150	0.148
	(0.005)**	(0.005)**
Observations	484308	484308

Standard errors in parentheses
* significant at 5%; ** significant at 1%
[a] the reference category is "Unassigned or assigned to an individual."
[b] The reference category is "US only, single state."

Table 6: Negative Binomial Estimation
Dependent Variable = Citations Received from Patents with US Inventors
Case Selection: US Patents with Foreign Inventors Issued 1990-1994

	(3)	(4)	(5)
Claims	0.004	0.004	0.004
	(0.000)**	(0.000)**	(0.000)**
Citations Made	0.013	0.012	0.012
	(0.000)**	(0.000)**	(0.000)**
Grant Year Dummies	(4, **)	(4, **)	(4, **)
Technology Class Dummies	(405, **)	(405, **)	(405, **)
Assignee Type[a]*:*			
US Firm	0.095	0.105	0.157
	(0.012)**	(0.012)**	(0.015)**
Non-US Firm	-0.062	-0.061	-0.038
	(0.008)**	(0.008)**	(0.010)**
US Government	0.178	0.167	0.333
	(0.088)*	(0.088)	(0.095)**
Non-US Government	-0.057	-0.038	-0.002
	(0.027)*	(0.027)	(0.030)
US University	0.272	0.280	0.329
	(0.044)**	(0.044)**	(0.053)**
US States (n)	0.296	0.293	0.287
	(0.011)**	(0.011)**	(0.012)**
English Language (1 = yes)	0.214		
	(0.007)**		
Legal System[a]*:*			
English		0.210	0.352
		(0.007)**	(0.019)**
French		-0.047	0.077
		(0.008)**	(0.015)**
Scandinavian		0.061	0.196
		(0.014)**	(0.019)**
Socialist		-0.179	0.129
		(0.046)**	(0.182)
Roman		0.089	0.394
		(0.048)	(0.055)**
Islamic		0.015	0.122
		(0.067)	(0.085)
Socio/Economic Indicators:			
Development (ln GPD/cap.)			0.103
			(0.017)**
Education (% Tert. Enroll.)			-0.000
			(0.000)
Scientific Publications			0.000
			(0.000)**
Foreign Direct Investment			0.002
			(0.005)
GP IPR Index			-0.067
			(0.018)**
Constant	-0.292	-0.287	-2.664
	(0.010)**	(0.011)**	(0.348)**
Observations	228732	228732	201029

Standard errors in parentheses
* significant at 5%; ** significant at 1%
[a] the reference category is "German."

CODING APPENDIX

We draw on Reynolds (2004) for our coding of the legal systems variables and create categorical variables based on origin of the system. We also cross-referenced our coding with La Porta et. al. (1999) who code significantly fewer countries than we do and only distinguish between common law systems and variants of civil law systems.

Reynolds distinguishes between four principal legal families. These are common law, civil law, socialist, and Islamic law. In addition he divides the civil law family into French, German, and Scandinavian. Moreover, he classifies a handful of countries as Roman-Dutch, which is a variant on a common law system. Dutch law is based on the French tradition (civil) but the Roman-Dutch law that evolves out of it is more similar to common law and this is the code that was transported to Botswana, Namibia, South Africa and Sri Lanka. Finally, Reynolds discusses less widespread indigenous categories that we have left out. For instance, we are not coding for African customary law which in many cases is important. Similarly we do not account for Confucian influence in Chinese and Korean legislation.

In a number of cases Reynolds articulates several different influences on the legal system of a country. The coding convention we followed was to code a country as belonging to a particular family if the influence of that system predominated. When the influence was decidedly mixed we coded the country as mixed (belonging to more than one category) and when the system was reported in transition we left the coding blank.

Coding in Western Europe, North America and Oceana is straightforward as countries there followed either the common law tradition or a variant on the civil law tradition. India is coded as a common law system despite the notable Islam/ Hindu influence because the religious law has been codified through the common law code. In Latin America the classification is generally very straight forward as the countries in question were mostly Spanish or Portugese, or British colonies, and adopted the legal tradition of the colonial country. An exception to this is the legal tradition in Suriname, which is Dutch in origin. Current Dutch legal code follows the civil law tradition and Suriname is, therefore, coded as such (Reynolds does not classify Suriname as Roman-Dutch). Furthermore, the legal system of Trinidad and Tobago was originally Spanish civil law, but changed to common law in 1802 when the Spanish formally ceded the island to Britain. Thus we code Trinidad and Tobago as a common law country. Also Nicaragua is classified as socialist from 1979-1990.

Generally we found that Islamic law is practiced with other types of law usually French but sometimes British and coded as such. For instance, Indonesia is classified as using Islamic law in addition to French civil law because "Islam (in this overwhelmingly Muslim country) is recognized … in the 1989 law on religious courts which restates and clarifies the position and role of Islamic law in the legal system." (Reynolds 2004). In other cases the Islamic influence did not seem sufficient to code as a significant. For example, In Ghana only the relatively small Muslim "population is governed by principles of Shari'a (the uncodified body of Islamic law and practices, based on the Holy Qur'an and the Sunnah of the Prophet) in regard to matrimonial matters." (Reynolds 2004) Similarly Guinea is not coded as having Islamic influence because according to Reynolds despite having a sizeable Muslim population Guinean legal code "closely adheres to the structure and principles of the French law". A few countries are coded as

exclusively Islamic including Oman and Malay.

A more unusual combination is Myanmar which we classified as both Common law and Socialist. In Hong Kong China's influence is recognized by adding Socialist coding in 1997 on. Romania is coded as using French civil law and Socialist law until 1990 because existing legal codes were not changed but piled upon with socialist legislation during the reign of Dej and later Ceausescu. Vietnam also follows both French and socialist legal tradition. Cameroon and Vanuatu merge British and French law and are coded as both. Albania from 1990 has incorporate both French and German civil law influences in its code.

We did simplify some very complex cases. For example, Congo is classified as French civil law despite Reynolds assertion that it is more complex than that and really there is no rule of law. Similarly, according to Reynolds the Philippines use a mixture of civil and common law. La Porta et. al. (1999), however, code the Philippines as a French civil law country and we follow their example. Taiwan is coded as a German civil law system despite the fact that the country is not independent.

Neither Reynolds (2004) nor La Porta et. al. (1999) have any information on San Marino but we code it as following French civil law because inhabitants are Italians and native San Marinos. There is no information on Swaziland either but it was a British colony and we code it as a Common law country.

Finally we have no information on Armenia, Azerbaijan, Yemen, Georgia, Kyrgistan after 1991, Iraq after 1990 (before 1990 it was Ottoman and socialist mixed with some common law), Nepal and Palau. Furthermore, according to Reynolds Bulgaria, and Hungary after 1990 are still developing a "western" legal system, and he notes that legal developments in Ukraine after 1991 are confusing. Therefore, we did not code these.