



WIPO Economics & Statistics Series

January

# 2014

## Economic Research Working Paper No. 15

Diaspora Networks, Knowledge Flows and Brain Drain

Ajay Agrawal



# Diaspora Networks, Knowledge Flows, and Brain Drain

Ajay Agrawal<sup>1</sup>

## Abstract

I summarize key findings from the literature on how distance, relationships, and ethnic ties influence knowledge flows and describe a model that relates emigration and the diaspora to knowledge flows. I then recap a key study that reports evidence of a link from the diaspora and knowledge flows to home country manufacturing productivity. Next, I summarize the ways in which intellectual property protection may influence knowledge flow patterns through incentives (market for ideas) and disincentives (anti-commons). Finally, I speculate on how diaspora knowledge flows and intellectual property may alleviate developing country low-productivity equilibria (“poverty traps”) caused by an underinvestment in specialized human capital.

## Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the views of the World Intellectual Property Organization or its member states.

## Acknowledgements

I am grateful for helpful feedback from Carsten Fink, Ernest Miguelez, Roberta Piermartini, Hillel Rapoport, and participants at the “WIPO Experts’ Meeting on Intellectual Property, the International Mobility of Knowledge Workers, and Brain Drain” hosted by the Economics and Statistics Division of the World Intellectual Property Organization in Geneva, Switzerland. Errors and omissions are my own.

---

<sup>1</sup> University of Toronto and NBER. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON, Canada M5S 3E6. [ajay.agrawal@rotman.utoronto.ca](mailto:ajay.agrawal@rotman.utoronto.ca).

## 1. Introduction

Economists are interested in understanding the determinants of knowledge flow patterns since knowledge is a primary input to innovation and innovation is central to economic growth. My focus here begins with one specific factor that influences knowledge flows: the diaspora. Members of the diaspora are important for three primary reasons: 1) they are costly to their home country since they represent lost human capital and localized knowledge spillovers, 2) they are valuable to their home country since they may send particularly high quality knowledge flows back, and 3) they may play a unique and critical role in solving the poverty trap problem once the home country reaches a certain level of development.

Given that the World Intellectual Property Organization is the setting for this presentation, it is incumbent on me to point out that I reference patents in two distinct contexts throughout this essay. First, patents are a common source of data in the empirical study of knowledge flows. In this case, my focus is not on the part patents play in facilitating intellectual property protection but rather the common role patent citations play as a proxy for knowledge flows. Second, I reference patents in their traditional role as a mechanism to confer property rights through enabling the owner to exclude others from use. I will be clear in distinguishing which is which in each case.

Perhaps the first large-sample study examining a determinant of knowledge flow patterns is the paper by Adam Jaffe, Manuel Trajtenberg, and Rebecca Henderson (1993, hereafter JTH), which examines the role of distance. It is perhaps not surprising that such an important advance was delivered by these particular authors since all three are students of Zvi Griliches, who laid the foundations for much of the empirical research on technological innovation and productivity over the past half century, beginning with his pioneering empirical study on diffusion utilizing the setting of hybrid corn (Griliches, 1958). The key methodological insight in the JTH paper is that patents contain three pieces of information that enable empirical research on knowledge flows at the idea level: 1) citations to prior art may be used as a proxy for knowledge flows, 2) geographic location of the inventors at the city level, and 3) classification of knowledge embodied in the claims of the patent. Together, this information enables the estimation of disproportionate knowledge flows between inventors at specific locations by employing a matching procedure to control for the underlying geographic distribution of ideas in a particular technology class. The primary finding reported in this paper is that knowledge flows are geographically localized, particularly at the MSA ("city") level.

Peter Thompson and Melanie Fox-Kean (2005) raise legitimate concerns with the method (questioning the ability of the matched pairs to fully control for the underlying geographic distribution of inventive activity in a narrowly defined class), which led to a debate published in the *American Economic Review* and ultimately resulted in scholars modifying elements of the method in subsequent studies but continuing to employ the general technique to estimate knowledge flows. Stefano Breschi and Francesco Lissoni (2001) also raise legitimate concerns regarding the interpretation of the JTH result as a measure of knowledge spillovers since, among other concerns, citations may reflect knowledge flows that are priced (e.g., licensing) and thus not externalities. I thus adopt the less incorrect term knowledge flows rather than spillovers here.

The JTH result raised the question of why. Why are knowledge flows geographically localized? Is it entirely due to communication costs? Co-location lowers the cost of face-to-face interactions often described as necessary for transferring tacit knowledge (Agrawal, 2006). Or might part of the localization finding be due to something more subtle, such as social relationships (perhaps originally facilitated by lower communication costs; see, Gaspar and Glaeser, 1998)?

Iain Cockburn, John McHale, and I (Agrawal et al, 2006) explore the relationship between co-location and knowledge flows by focusing on individuals who are originally co-located and then move away. Our main finding is that when an inventor creates an invention (proxied by being granted a patent) in their new location, it is cited disproportionately by inventors from their former location. We interpret this as evidence of the enduring social capital hypothesis – that knowledge flows are influenced by social capital that may be formed through co-location but endures even when individuals are no longer co-located (and no longer enjoy low-cost communications, such as face-to-face interactions). Furthermore, we find the knowledge flow premium associated with prior co-location is particularly strong for across-field knowledge flows, where arguably social capital is relatively more important. Stefano Breschi and Francesco Lissoni (2009) then explore a more nuanced view of social relationships and measure knowledge flows between social networks established from co-invention. They report that after controlling for the co-location network, the residual effect of geographic proximity on location is greatly reduced. In other words, they interpret their result as implying that the geographic localization of knowledge flows is due to the localization of relationships.

Devesh Kapur, John McHale, and I (Agrawal et al, 2008) then examine whether other forms of social capital have a similar effect on knowledge flows. In particular, we focus on the social capital associated with co-ethnicity. Again using a derivative of the methodological technique devised by JTH and identifying co-ethnic inventors using name analysis (Indian inventors living in the US and Canada), we report that co-location and co-ethnicity both predict knowledge flows. Importantly, however, they are substitutes. In other words, the marginal benefit of co-location is significantly less for co-ethnic inventors. We interpret this finding as providing further evidence that social capital is a key determinant of knowledge flows. Co-location is one way to generate social capital between individuals; however, there are other ways, such as shared ethnicity.

This brings us to the diaspora. Devesh Kapur, John McHale, Alex Oettl, and I (Agrawal et al, 2011) bring together the prior findings on co-location and co-ethnicity to consider how emigration may influence knowledge access by inventors in developing countries. We develop a model that considers three factors associated with the diaspora: 1) the human capital that is lost when an inventor emigrates, 2) the localized knowledge flows that are lost when an inventor emigrates, and 3) the potentially higher quality knowledge that flows back from the diaspora to inventors who remain in the home country (“co-national”). We then use a derivative of the JTH technique to estimate the knowledge flow parameters of the model. Since the model addresses the central issue of this paper, I reproduce it in the following section with the permission of the original publisher.

## 2. A Model of Diaspora Knowledge Flows

### 2.1 Permanent migration

We introduce the concept of a Knowledge Flow Production Function (KFPF), the probability of a domestic innovator receiving knowledge from any other innovator based on structural aspects of their relationship, to develop a simple model of an optimal innovator diaspora. Our focus is on knowledge production in a relatively poor country, which we call India without loss of generality. The essential idea is that the productivity of India-residing innovators depends on their access to knowledge. This access in turn depends on their relationships to other innovators and also on the productivity of those innovators. We allow connectivity to be affected by co-location and co-nationality and also for the possibility that innovators are more productive abroad because of better incentive structures and resources (Kahn and McGarvie, 2011). The emigration of an innovator results in a direct loss to the stock of Indian innovators, thinning domestic knowledge networks, but could actually increase total knowledge access if the diasporic linkages and productivity gains are large enough. The model's goal is to identify the size of the diaspora that maximizes the access to knowledge of India-residing innovators.<sup>2</sup> The KFPF captures the probability of a knowledge flow between any pair of innovators (at least one of whom is a resident of India) based on certain structural relationships between those innovators. We express the probability of a knowledge flow to a particular Indian innovator,  $i$ , from another innovator,  $j$ , as:

$$(1) \quad K_{ij} = f + \alpha_{ij}\mathcal{M} + \beta_{ij}\mathcal{D},$$

where  $f$  is the (base-case) probability of a knowledge flow if the other innovator is neither a resident of India nor a member of the Indian diaspora,  $\alpha_{ij}$  is a dummy variable that takes the value of 1 if innovator  $j$  is also a resident of India,  $\gamma$  is the proportionate knowledge-flow premium from being co-located,  $\beta_{ij}$  is a dummy variable that takes the value of 1 if  $j$  is a member of the Indian Diaspora, and  $\delta$  is the proportionate premium for being in the diaspora. Note that the value of  $\gamma$  reflects the combined effects of co-location and the (possibly negative) relative productivity effect of doing science in India, whereas the value of  $\delta$  reflects the effect of the diaspora connection and any productivity gap that might exist between members of the diaspora and foreigners. Denoting the total number of Indian innovators (both India-based and emigrant) as  $N$ , the total size of the Indian scientific diaspora as  $D$ , and the total number of foreign innovators as  $Z$ , we express the total (expected) knowledge flow to  $i$  with this knowledge access equation:

$$(2) \quad K_i = Zf + (N - D - 1)(1 + \gamma)f + D(1 + \delta)f.$$

We then find the aggregate knowledge access of India-residing innovators by multiplying both sides of (2) by the total number of such innovators:

---

<sup>2</sup> The model allows for a trade-off between the costs of weakened local knowledge networks and the benefits of access to more distant knowledge. Papers in urban economics have highlighted other potential trade-offs associated with labor pooling. Combes and Duranton (2006) develop a model in which labor pooling has two opposing effects: It allows greater access to knowledge produced by other firms, but the potential for one's own workers to be poached forces firms to pay higher wages to retain their workforce. Gerlach et al. (2009) develop a model with the same deglomerative force but in which the agglomerative force comes from asymmetric R&D investments that produce a diversified portfolio of technologies at the industry level.

$$(3) \quad K = (N - D)K_i = (N - D)Zf + (N - D)(N - D - 1)(1 + \gamma)f + (N - D)D(1 + \delta)f.$$

We assume innovation depends on both the access to knowledge and the absorptive capacity to turn that knowledge into valuable economic output. In this paper, we focus only on knowledge access and assume it is positively associated with output:

$$I_i = I(K_i) \quad ; \quad \frac{\partial I_i}{\partial K_i} > 0. \text{ Of course, the knowledge access to innovation will be}$$

country specific and depend, *inter alia*, on the available capital stock, the presence of complementary human capital, and the security of property rights.

We find the diaspora size,  $D^*$ , that maximizes national knowledge access (and thus innovation) from the first-order condition:

$$(4) \quad \frac{\partial K}{\partial D} = 2D^*(\gamma - \delta) - Z - N(1 + 2\gamma - \delta) + (1 + \gamma) = 0.$$

Rearranging Equation (4), we obtain an expression for the optimal diaspora as a fraction of the total stock of Indian innovators:

$$(5) \quad \frac{D^*}{N} = \left( \frac{1 + 2\gamma - \delta}{2(\gamma - \delta)} \right) + \left( \frac{1}{2(\gamma - \delta)} \right) \left( \frac{Z}{N} \right) - \left( \frac{1 + \gamma}{2(\gamma - \delta)} \right) \left( \frac{1}{N} \right).$$

Equations (3) through (5) allow us to characterize the conditions under which a diaspora is beneficial for knowledge access and innovation. We do this in two steps. First, an examination of Equations (3) and (4) reveals that, for this first-order condition to identify a maximum, we require from the second-order condition that  $\delta$  is greater than  $\gamma$ :

$$(6) \quad \frac{\partial^2 K}{\partial D^2} = 2(\gamma - \delta) < 0 \quad \Rightarrow \quad \delta > \gamma.$$

Otherwise, the national knowledge access will decline monotonically with the size of the diaspora (see the first equality in Equation (4)). We first assume that this condition does not hold. A positive diaspora is never beneficial in this case. We can give this necessary condition a more intuitive explanation. Suppose in the extreme that the potential emigrants contribute nothing directly to domestic innovation while at home. Their only contribution comes indirectly from the knowledge that flows from them to other domestic innovators. Whether their absence helps or harms in that case depends simply on whether domestic innovators access more knowledge from them when at home or abroad, i.e., on the relative magnitudes of  $\delta$  and  $\gamma$ .

Second, we use Equation (7) to identify the necessary and sufficient condition for a strictly positive diaspora to be beneficial:

$$(7) \quad \frac{D^*}{N} > 0 \quad \Leftrightarrow \quad \delta > 1 + 2\gamma + \frac{Z}{N} - \frac{1 + \gamma}{N}.$$

This condition is quite stringent. Even in the extreme case where  $N$  is sufficiently large enough that we can ignore the last two terms and where there is no co-location premium (i.e.,  $\gamma = 0$ ), the diaspora premium must be greater than 100% for a diaspora to be beneficial for the total knowledge flow to India-residing innovators.

From Equation (5), we can see that the optimal diaspora share converges to one half as  $\delta$  approaches infinity. In other words, it will never be optimal for a country to have more than half its innovators abroad. Although in reality we expect the optimal diaspora share to be well below one half, this finding is of interest because there are several countries for which the number of tertiary-educated nationals residing abroad is greater than the number residing at home (Docquier and Marfouk, 2005). These general emigrant shares are likely to underestimate the share of innovators, given the tendency for emigrant shares from poor countries to rise with education level. The model suggests that this is detrimental to knowledge production no matter how large the productivity gains are from emigrating and no matter how strong the diasporic connections. This result implies that countries must have a sufficient number of innovators at home to reap the benefits of emigrant-related productivity gains and diasporic connections.

## 2.2 Circulatory migration

The model with permanent migration abstracts from one potentially important element: the return of emigrant innovators. Such returnees are likely to have developed connections with foreign innovators while away, connections that may endure on their return to facilitate ongoing knowledge flows.<sup>3</sup> To explore the implications of return, we next examine the steady state of a simple extension of the model that allows for circulation.

At any point in time, the change in the diaspora share mechanically depends on the emigration rate ( $e$ ), the return rate ( $r$ ), the growth rate of new Indian scientists ( $n$ ), and the initial diaspora share:<sup>4</sup>

$$\begin{aligned}
 d\left(\frac{D}{N}\right) &= \frac{1}{N} dD - \frac{D}{N^2} dN \\
 (8) \qquad &= \frac{1}{N} (e(N - D) - rD) - \frac{D}{N} n \\
 &= e - (e + r + n) \frac{D}{N}.
 \end{aligned}$$

---

<sup>3</sup> Agrawal, Cockburn, and McHale (2006) provide evidence of the impact of enduring social capital acquired during past co-location on subsequent knowledge flows.

<sup>4</sup> The emigration rate is the fraction of the stock of India-residing innovators ( $N - D$ ) who emigrates each period, the return rate is the fraction of the innovator diaspora ( $D$ ) who returns each period, and the new innovator growth rate is the proportionate growth in the total stock of Indian innovators ( $N$ ).

Setting Equation (8) equal to zero, we have an expression for the *steady-state* diaspora share:

$$(9) \quad \left( \frac{D}{N} \right)^{ss} = \frac{e}{e + r + n}.$$

For a given steady-state diaspora share and a given  $n$ , the steady state is consistent with an infinite number of  $(e, r)$  pairs. One possibility is that a given diaspora share is observed with very low emigration and return rates, such that the diaspora and the stock of scientists remaining in India have the character of “stagnant pools.” However, we can observe the same diaspora share with much higher emigration and return rates, such that the diaspora and India-residing stocks have more the character of “circulating pools,” innovators whom Saxenian (2006) calls the “New Argonauts” after the Greeks who sailed with Jason in search of the Golden Fleece. The nature of the India-residing stock is likely to have implications for the strength of their connections to domestic, diasporic, and foreign scientists, with the relative strength of connections to innovators abroad increasing with the propensity to circulate.

Given perpetual circulation, the expected fraction of time that any Indian innovator will spend in the diaspora will converge to the steady-state diaspora share for any strictly positive return rate. Looked at from the viewpoint of innovators currently residing in India, the expected fraction of time spent abroad in the past is therefore increasing in the steady-state diaspora share. An implication is that with a positive return rate, a higher diaspora share is likely to be associated with stronger connections to foreign innovators.<sup>5</sup> This suggests a potential problem with inferences about optimal diaspora size based on the static model. We develop the static model on the premise of proportional co-location and diaspora premiums that are independent of the size of the diaspora itself. This independence allows us to estimate these premiums and then make inferences about the optimal size of the diaspora. However, if a larger diaspora share is associated with stronger connections to innovators abroad, then it is likely that the size of the diaspora will affect the proportional co-location and diaspora premiums. But when these premiums depend on the size of the diaspora, we face the problem that we cannot use estimates of these premiums (based on a time period with a given diaspora) to infer the size of the optimal diaspora. We outline our method for identifying the importance of return in the empirical strategy section below.

### 2.3 Heterogeneous innovators and non-random selection

We have assumed that all innovators are equally productive. However, we can weaken this assumption without affecting the results if we assume that emigrants and returnees are random selections from the stocks of India-residing innovators and the diaspora, respectively. The results are obviously affected, however, if emigrants and returnees are non-random selections from their respective pools. Suppose, for example, that the most productive innovators have a higher probability of emigrating (possibly because they have a higher probability of qualifying for a visa such as the U.S. H-1B).

---

<sup>5</sup> When the return rate is zero, such that the current India-residing stock has spent no time abroad, the strength of the connection to foreign scientists is independent of the size of the diaspora.



This positive selection will tend to augment the absence-related loss to India, suggesting an even lower optimal diaspora. Suppose further that returnees are a positive selection of the already positively selected diaspora. It is possible that a few truly outstanding returnees, coming back with significantly enhanced productivity due to their time spent abroad, could have a major impact on Indian innovation. In this case, our model would give a misleading picture of the long-run effect of migration.

#### 2.4 Knowledge access and the value of an innovation

A core idea of the model is that knowledge access drives innovation. To keep the model as simple as possible, we have made the restrictive assumption that the way relationships facilitate knowledge access is the same for all innovators. One obvious concern is that the KFPF differs systematically based on the value of the innovation. For example, high-value innovations may draw relatively more on frontier knowledge through the diaspora. As another example of how the KFPF may be context specific, Nanda and Khanna (2010) find that diasporic connections are more important for Indian software entrepreneurs operating in weak institutional environments.

### **3. The Diaspora and Knowledge Flows: Empirical Evidence**

We use patent citation data and a derivation of the JTH technique to estimate the knowledge flow parameters in the above model. The estimated co-location premium (disproportionate level of knowledge flows from other inventors in India to a given Indian inventor) is significantly larger than the estimated diaspora premium (disproportionate level of flows from Indian diaspora back to a given inventor in India) which, interpreted through the lens of the model, implies that the net effect of emigration on domestic innovation is negative. The benefits from the diaspora, in terms of knowledge access, do not compensate for the loss of localized flows, let alone the loss of human capital. However, we also report a caveat to this finding. As we restrict the sample of Indian inventions to those of increasing importance (as measured by the number of citations they receive from subsequent patents), the relative role of diaspora knowledge flows increases sharply while at the same time the importance of local flows declines. Narrowing the sample to only the 93rd percentile and above, we see a large diaspora effect (almost 10 times the magnitude as that for the overall sample), while the co-location effect is about four times smaller and no longer statistically significant. This caveat is important since it is well known that the value of innovations increases nonlinearly with the number of citations (Trajtenberg, 1990).

We also examine the relative quality of emigrants and returnees. We find very little difference in the quality of returnees versus non-returnees (conditional on returnees continuing to invent upon returning, since we measure quality by way of forward citations to patents). However, we find evidence that emigrants are highly positively selected. Inventors who will subsequently emigrate receive, on average, about nine times as many citations as those who do not emigrate. This finding reinforces the inference based on the simple model above; inventor emigration harms knowledge access and domestic innovation (notwithstanding the caveat regarding the most impactful inventions). Bill Kerr (2008) provides evidence that not only links diaspora to knowledge flows but importantly also establishes a link to manufacturing productivity in the home country. He begins this study by reporting diaspora knowledge flows using patent citation data and inventor name ethnicities. The paper reports that own-ethnicity citations are 50% higher than citations to other ethnicities.

Essentially, this result implies that an invention by a US-based inventor with a Chinese name receives 50% more citations from Chinese inventors in China than would be expected given the distribution of inventions in that technology class across ethnicity space. In a separate study (Kerr, 2010), he demonstrates that this correlation can be reduced to 20-30% using stricter controls for technology classification, implying ethnic clustering by technology. He further shows the co-ethnicity knowledge flow premium peaks at approximately five years following the invention, followed by a gradual decline over the next five years.

Perhaps most importantly, he then goes on to report a link between the diaspora and manufacturing productivity in the home country using the Industrial Statistics Database of the UN's Industrial Development Organization. He regresses productivity (manufacturing output per employee) on the ethnic human capital stock by industry and country using panel data with country and industry fixed effects such that the estimation is driven from within industry variation in diaspora size and home country output. He reports that a 10% increase in US ethnic research is associated with a 1% increase in foreign output. He decomposes this result (labor productivity gains versus expansion in employment) and finds that labor productivity growth facilitates most of the gains. Importantly, given the potential for omitted variable bias, he shows that this result is robust to a shock to diaspora size caused by a surge in immigration of scientists and engineers due to a revision to the US quota system. This is the most compelling evidence to date linking the innovative diaspora to home country manufacturing productivity.

#### **4. Intellectual Property and Knowledge Flows**

We now turn to considering the role of intellectual property protection on knowledge flows. The role of patents in creating incentives for innovators to innovate by conferring monopoly rights over ideas (by way of excluding others from use) in exchange for disclosure is well known. However, our focus here is on the role of patents in facilitating knowledge flows as opposed to knowledge creation. On the one hand, patents enhance knowledge flows in two ways. First, they promote disclosure. Second, they facilitate trade. On the other hand, they may also inhibit knowledge flows due to the anti-commons effect.

The role of patents in promoting disclosure is obvious and a feature of all major patenting systems. Teece (1986) and Gans and Stern (2003) describe the role of patents in facilitating trade in the market for ideas. The specific role of patents in enhancing trade is perhaps best illustrated in Gans, Hsu, and Stern (2008), where they estimate the effect of the patent allowance date on the timing of licensing activity. The Notice of Patent Allowance event is the date on which the US Patent and Trademark Office announces to the applicant the rights the Office will grant to them. They argue that if the market for technology licenses is efficient, then the timing of licensing should be independent of whether patents have already been granted. However, if there are imperfections in the market for ideas due to, for example, information asymmetries, search costs, or a need to disclose complementary unprotected knowledge, then formal IP rights may be important for facilitating gains from trade.

To estimate the role of formal IP rights on trade, the authors exploit the variation in patent allowance and licensing lags across inventions and the timing of licensing relative to patent allowance. Using data from 198 patent-license pairs (i.e., the sample is conditioned on inventions that were licensed), they estimate that patent allowance increases the hazard rate of achieving a licensing agreement considerably - by 70-80%. The authors also find that the role of patents is less important for facilitating trade in settings where alternative institutions exist (copyright protection in software, reputation preservation in Silicon Valley). Overall, this paper provides compelling evidence that patents play a very significant role in enhancing knowledge flows via trade (or at least knowledge access).

At the same time, a growing literature has focused on the negative effects of patents on knowledge flows – the so called “anti-commons” effect. Two recent empirical papers by Murray and Stern (2007) and Williams (2013) have established the frontier on estimating an IP penalty on knowledge flows. They each examine the role of IP in the context of cumulative innovation: Do IP rights on existing knowledge hinder subsequent innovation? They also both estimate a penalty on knowledge flows due to patenting knowledge relative to a counterfactual where knowledge is created and not protected but rather made freely available in the public domain. Murray and Stern estimate a 10-20% IP penalty and Williams a 20-30% IP penalty in their respective empirical settings. The main empirical challenge to estimating an IP penalty is identification. Some knowledge is protected while other knowledge is not. The decision of whether or not to protect knowledge with patents is not random. Therefore, simply comparing the knowledge flows associated with patented inventions versus those associated with inventions that are not patented but rather freely available is subject to bias. For example, perhaps knowledge that is more valuable is more likely to be patented. If so, then comparing knowledge flows between patented and non-patented knowledge would underestimate the IP penalty since the comparison does not take into account that the patented knowledge is of higher quality and thus likely to receive more citations. In other words, any comparison must control for differences in the quality of the idea and its propensity to generate knowledge flows.

The aforementioned papers stand out in terms of the creative approach each employs to address the identification challenge. Murray and Stern exploit an insight that occasionally ideas are captured as “dual knowledge,” such that they are simultaneously patented and published in publicly available journals. In other words, they compare citations associated with the same piece of knowledge under two regimes (pre- versus post-IP protection), providing a comparison with a counterfactual that arguably perfectly controls for the quality of the idea and its propensity to generate knowledge flows since it is the same piece of knowledge under both regimes. Williams exploits an insight that in the race to sequence the human genome, a private, for-profit effort occurred simultaneously along with a public effort such that some gene discoveries were patented and temporarily kept secret while others were published and made freely available in the public domain. Once again the author is able to, in some sense, perfectly control for the quality of the knowledge and its tendency to generate knowledge flows since the same piece of knowledge exists first in the IP-protected regime and subsequently in the public domain. Furthermore, other comparable pieces of knowledge (other genes) are also used as counterfactuals to compare knowledge flows from ideas that were temporarily protected to those that were never protected. I describe the main features of each study below.

Murray and Stern employ the scientific paper as their unit of analysis. They construct a sample of 340 papers published in Nature Biotechnology during the period 1997-1999, of which 169 have associated patents (i.e., patent-paper pairs). They use a count of the citations received by the focal paper as their dependent variable and compare the citations received by “treated” papers (where the knowledge is also patented) to citations received by control papers (no patents) both before and after treatment (the point in time where the patent is issued to the knowledge embodied in the patent-paper pair), exploiting the patent grant delay. In other words, they employ a difference-in-differences type estimation. They also estimate a single difference (before versus after) on patent-paper pairs using paper fixed effects, estimating the average within knowledge-piece variation pre- versus post-IP protection relative to the trend in citation rates for papers with similar characteristics. They report three main findings: 1) patented knowledge receives more citations than non-patented knowledge, on average, but this seems mostly due to location and the number of authors rather than unobserved quality, 2) the citation rate declines by 10-20% per year after a patent is granted (IP penalty), and 3) the IP penalty is stronger for public sector coauthors, such as university researchers. Williams employs the gene as her unit of analysis. She constructs her sample using 27,882 genes, 6% of which were discovered by the private company Celera. She employs three main dependent variables: 1) publications investigating genotype-phenotype links, 2) knowledge about genotype-phenotype links, and 3) gene-based diagnostic tests. She then employs two main empirical tests. First, she examines the within-gene variation in subsequent innovation relative to the IP regime the gene is in (during versus after IP protection). Second, she examines the link between the duration during which the gene was under Celera IP and the level of subsequent innovation. She interprets her results thusly: “[I]f Celera genes had counterfactually had the same rate of subsequent innovation as non-Celera genes, there would have been 1,400 additional publications between 2001 and 2009, and 40 additional diagnostic tests as of 2009.”

Although these two papers offer valuable contributions to our understanding of the potential costs of IP to cumulative innovation, they both leave two first-order questions unanswered. First, both papers are silent on the positive effects of IP in terms of creating incentives for innovation. For example, the Murray and Stern paper notes that patented papers receive more citations than non-patented papers, on average. While that could simply be a correlation driven by the unobserved quality of ideas (higher quality ideas are likely to attract more citations and are also more likely to be patented), it could also be causal (patents create the incentive to develop higher quality ideas). Second, neither paper sheds light on the mechanism through which IP causes the estimated decline in knowledge flows. Williams speculates on three potential mechanisms through which the IP penalty may occur: 1) asymmetric information between the original inventor and the potential subsequent inventor, leading to bargaining frictions (contracting between inventors is necessary due to a “scarcity of ideas” on the part of the original inventor, who can’t imagine all the potential follow-on inventions); 2) disclosure problem (also known as “Arrow’s paradox”), where the value of the idea is compromised by sharing it with the potential buyer due to potential imitation, but the buyer needs to know the details of the idea in order to properly assess its value; and 3) transaction costs associated with uncertainty over the academic research exemption. Understanding the relative role of these and other transaction costs in generating the IP penalty is important because they offer insight into how labor mobility and IP will influence knowledge flows to developing economies.

For example, if the disclosure problem is a central cause of the IP penalty and if this problem is reduced through personal relationships between inventors, then labor mobility will increase the propensity of knowledge flows across national boundaries due to relationship formation between the diaspora and other co-located inventors. Similarly, if information asymmetry or uncertainty over the academic exemption is a central cause of the IP penalty, then knowledge flows across borders may be significantly enhanced as ideas are presented in more standardized formats online, even without extensive labor mobility. Both of these topics - innovation incentives from IP and the mechanisms underlying the IP penalty - represent first-order research questions that would provide a significant contribution to the literature.

## **5. Knowledge Flows, Diaspora, and the Poverty Trap**

Ben Jones (2008) proposes a theory predicated on human capital investments to explain poverty traps. His central innovation is characterizing human capital investments in two dimensions, quantity and quality, rather than just the former. Quantity refers to the traditional human capital metric concerning the duration of education, while quality refers to the degree of specialization. Individuals may invest in generalist or specialist skills. The benefit of specialist skills is that they generate higher collective productivity when combined with other specialists than generalist skills. The cost of specialist skills is that they generate lower productivity when they are not combined with other specialists. Poverty traps arise in economies with a thin initial stock of specialists, since individuals will refrain from investing in specialist skills since the market for co-specialists is thin. This will lead to an increasingly thin market for specialists, relative to economies with thick markets.

Diasporas may alleviate poverty traps in three ways. First, the prospect of leaving the poor country may provide the otherwise missing incentives for locals to invest in specialized human capital skills (Docquier and Rapoport, 2012). Not all who make such an investment actually leave. To the extent those who remain in the developing economy are able to apply their skills effectively with others, the prospect of emigration may help to solve the coordination problem.

Second, individuals who do emigrate, join the diaspora, and acquire specialized skills in an environment that provides high returns to such skills thus motivating investment in human capital, may one day return to their home country (Boeri et al, 2012). To the extent that the flow of returnees is above a critical threshold, emigration may provide a pathway for developing countries to break out of the low-productivity equilibrium, providing incentives for their citizens to invest in specialized skills. This at first may be in order to emigrate, then to collaborate with returnees, and eventually to coordinate with other domestic workers who also have invested in specialized skills.

Finally, and most importantly for this paper, the diaspora may provide access to knowledge to individuals in the developing economy. These knowledge flows themselves may increase the returns to specialization under the condition that the domestic workforce is sufficiently skilled to enjoy reasonable returns from utilizing these knowledge flows. In other words, knowledge flows from the diaspora may play an important role in reducing poverty trap effects that otherwise discourage locals from investing in specialized human capital development.

To the extent that formal intellectual property protection, through patents for example, facilitates trade as described above (Gans et al, 2008) and that the trade in ideas is disproportionately likely between diaspora and the home country as described in Agrawal et al (2008) and Kerr (2008), then it is plausible that patents and the diaspora are complements with respect to knowledge flows to the home country. In that case, a well-functioning patent system may amplify the benefit of a growing skilled diaspora in a highly productive country like the US (Kahn and MacGarvie, 2013). In other words, these two factors might be particularly potent in combination and could explain the especially strong diaspora effect for high technology as well as for China as reported in Kerr (2008). Alternatively, increased intellectual property protection could diminish knowledge flows from diaspora to individuals in their home country. For example, in the context of pharmaceuticals, the effect of increasing the efficacy of intellectual property protection (in either the United States or South Africa) on knowledge flows from the South African diaspora in the US to individuals “back home” in South Africa could be negative, rather than positive as suggested above. Perhaps the interaction of IP protection and diaspora knowledge flows depends on certain economic properties of the product? Whether they are complements in terms of alleviating the poverty trap is an empirical question. These topics are at the frontier of this literature and offer potentially fruitful directions for future research.

## **6. Conclusions**

The most significant shift in perspective on knowledge flows over the past 20 years (since the publication of JTH), has been in the emphasis on the role of social rather than geographic distance as a determinant of knowledge flow patterns. While spatial agglomeration remains a central topic among both scholars and policy makers, researchers have increasingly recognized the importance of social relationships and networks as the primary mechanism underlying this phenomenon. Since the diaspora is perhaps the most potent force to establish social relationships between high-income and lower-income nations, its members necessarily play an important role in shaping the flow of knowledge between these regions.

Still, many important questions remain unanswered. How does the shifting of intellectual property regimes in developing economies, in response to TRIPS for example, influence the migration of skilled workers and knowledge flow patterns? Furthermore, why are the diaspora from some countries more effective than others at facilitating knowledge flows back to their home countries (e.g., China vs. India in Kerr [2008])? Moreover, to what extent are strictly enforced patent systems in developing countries a complement to active diaspora abroad for generating north-south knowledge flows?

Given the importance of knowledge diffusion for productivity growth, particularly for low-income countries, insights into the microfoundations – individual costs and benefits that influence knowledge flow patterns – are not only interesting but also important. Enhancing our understanding of the microfoundations of knowledge flows will improve our ability to understand economic growth. This comprehension is also a prerequisite to effectively setting policy or designing strategy to influence knowledge flow patterns. Progress continues on this front. The number of scholars working with patent and publication data to study knowledge flows continues to grow. As a result, the quantity of knowledge flow-related research tools is increasing, and the quality of data continues to improve.

Furthermore, other mechanisms for measuring the flow of ideas (e.g., online social media text, which leaves a data trail on communication and idea exchange) are opening new avenues for research.

It has been 55 years since Zvi Griliches (1958) reported key insights into the diffusion patterns of hybrid corn, and we still have much to learn.

## References

- Agrawal, Ajay** (2006) "Engaging the Inventor: Exploring Licensing Strategies for University Inventions and the Role of Latent Knowledge" *Strategic Management Journal*, 27, 63-79.
- Agrawal, Ajay, Iain Cockburn, and John McHale** (2006) "Gone But Not Forgotten: Labor Flows, Knowledge Spillovers and Enduring Social Capital" *Journal of Economic Geography*, 6(5), 571-591.
- Agrawal, Ajay, Devesh Kapur, and John McHale** (2008) "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data" *Journal of Urban Economics*, 64, 258–269.
- Agrawal, Ajay, Devesh Kapur, John McHale, and Alex Oettl** (2011) "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation," *Journal of Urban Economics*, 69(1), 43-55.
- Boeri, Tito, Herbert Bruecker, Frederic Docquier, and Hillel Rapoport (eds)** (2012), Brain drain or brain gain? The international competition to attract high-skill migrants, Oxford University Press.
- Breschi, Stefano and Francesco Lissoni** (2009) "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows," *Journal of Economic Geography*, 9, 439-468.
- Breschi, Stefano and Francesco Lissoni** (2001) "Knowledge spillovers and local innovation systems: a critical survey," *Industrial and Corporate Change*, 10(4), 975-1005.
- Combes, P., Duranton, G.** (2006) "Labour pooling, labour poaching and spatial clustering," *Regional Science and Urban Economics* 36, 1–28.
- Docquier, F. and A. Marfouk** (2005) International migration by educational attainment. In: Özden, C., Schiff, M. (Eds.), International Migration, Remittances, and the Brain Drain. The World Bank and Palgrave Macmillan, Washington DC and New York, NY.
- Docquier, F. and H. Rapoport** (2012) "Globalization, brain drain, and development" *Journal of Economic Literature*, 50(3): 681-730.
- Gans, Joshua and Scott Stern** (2003), "The Product Market and the Market for Ideas: Commercialization Strategies for Entrepreneurs," *Research Policy*, 32 (2), 333-350.
- Gans, Joshua, David Hsu, and Scott Stern** (2008), "The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays," *Management Science*, 54(5), 982-997.
- Gasper, Jess, and Edward L. Glaeser** (1998), "Information Technology and the Future of Cities," *Journal of Urban Economics*, 43(1), 136-156.
- Gerlach, H., Ronde, T., Stahl, K.** (2009) "Labour pooling in R&D intensive industries," *Journal of Urban Economics* 65, 99–111.
- Griliches, Zvi** (1958) "Research Costs and Social Returns: Hybrid Corn and Related Innovations." *Journal of Political Economy*, 66(5), 419-431.
- Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson** (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, August, 577-598.
- Jones, Benjamin** (2008), "The Knowledge Trap: Human Capital and Development Reconsidered," National Bureau of Economic Research Working Paper 14138.
- Kahn, Shulamit and Megan MacGarvie** (2013) "How Important Is US Location for Research in Science?" *Review of Economics and Statistics*, forthcoming.
- Kerr, William** (2008), "Ethnic scientific communities and international technology diffusion," *The Review of Economics and Statistics* 90 (3), 518–537.



- Kerr, William** (2010), "The Agglomeration of U.S. Ethnic Inventors," in ed. Edward L. Glaeser Agglomeration Economics. The University of Chicago Press.
- Murray, Fiona and Scott Stern** (2007), "Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis," *Journal of Economic Behavior and Organization*, 63(4), 648-687.
- Nanda, Ramana, and Tarun Khanna** (2010) "Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry," *Journal of Economics and Management Strategy*, 19(4), 991-1012.
- Saxenian, A.** (2006) The New Argonauts: Regional Advantage in a Global Economy. Harvard University Press, Cambridge, MA.
- Teece, David** (1986) "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing, and Public Policy," *Research Policy*, 15(6), 285-305.
- Thompson, Peter and Melanie Fox-Kean** (2005) "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment" *American Economic Review*, 95(1), 450-460.
- Trajtenberg, Manuel** (1990) "A penny for your quotes: patent citations and the value of innovations" *RAND Journal of Economics* 21 (1), 172–187.
- Williams, Heidi** (2013), "Intellectual property rights and innovation: Evidence from the Human Genome," *Journal of Political Economy*, 121(1), forthcoming.