Does Scientist Immigration Harm US Science? An Examination of Spillovers*

Ajay Agrawal
University of Toronto and NBER
ajay.agrawal@rotman.utoronto.ca

John McHale
National University of Ireland, Galway
john.mchale@nuigalway.ie

Alexander Oettl
Georgia Institute of Technology
alex.oettl@scheller.gatech.edu

April 18, 2017

Abstract

The recruitment of foreign scientists enhances US science through an expanded workforce, but potentially could also cause harm by displacing better connected domestic scientists, thereby reducing localized knowledge spillovers. We develop a model in which a sufficient condition for the absence of overall harm is that foreign recruits are equally well connected to US scientists as the domestic scientists they displace. To test this condition, we conduct a hypothetical experiment in which each immigrant displaces an appropriately matched domestic scientist. Our measure of connection is subsequent citations to the scientist's work by other US scientists. Although we find that prior to their move immigrants are significantly less connected to US scientists than their matches, the post-immigration catch up in connection patterns is rapid. Once in the US, immigrant scientists are cited by US scientists at rates that are at least as great as their domestic matches. We find that the immigrant forward citation premium tends to be greater where they are relatively isolated from co-nationals and also where they come from countries where the use of English is common. Overall, we do not find evidence that immigrant scientists harm US science through a crowding out of better connected US scientists.

JEL Classifications: J61, O31, O33 **Keywords:** immigration, displacement, spillovers, knowledge flows, scientists, externalities

*We thank William Lincoln and Mercedes Delgado for helpful comments. This research was funded by the Centre for Innovation and Entrepreneurship at the Rotman School of Management, University of Toronto, the Martin Prosperity Institute, the Social Sciences and Humanities Research Council of Canada, and the Kauffman Junior Faculty Fellowship. All errors are our own.

1 Introduction

A large literature documents the importance of knowledge spillovers to the advancement of science, much of which are facilitated by intricate networks of co-located and non-co-located peers (Jaffe et al., 1993; Waldinger, 2010; Azoulay et al., 2010). In recent decades, US Science has also become increasingly internationalized, with rapid growth in the number of foreign-born scientists and engineers (Stephan, 2012; Freeman et al., 2015). Between 2003 and 2013, the number of immigrant scientists increased from 3.4 million to 5.2 million (Lan et al., 2015). For "Physical and Related Scientists," the number of immigrant scientists increased by 17,000 while the number of US-born scientists actually decreased by 14,000. With a downward sloping demand curve for scientists and an upward sloping supply curve for US-born scientists, standard market analysis predicts that there will be displacement of US-born scientists (Borjas, 2007; Borjas and Doran, 2012).

A central theme of the economics of immigration literature has been the measurement of wage and employment displacement effects (Borjas, 2005; Kerr and Kerr, 2011; Peri, 2012; National Academies of Sciences, Engineering, and Medicine, 2016). A large body of work has also explored the aggregate productivity effects of immigration. In the 'canonical model' (see, e.g., Borjas, 2014), the existence of aggregate gains from immigration depend on the displacement of native workers. The relatively small aggregate gain implied by this model has led researchers to look for evidence of externalities, especially in the form of knowledge spillovers (Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Peri, 2012; Peri et al., 2013). This has in turn led to an emphasis on peer networks that support knowledge exchange, work that connects to the large body of evidence that documents the importance of local knowledge spillovers (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Agrawal et al., 2006; Oettl, 2012). But if local networks are critical to knowledge exchange within US science, an inflow of immigrants that displaces native workers could disrupt local knowledge networks if the immigrants are less connected to US science than the domestic scientists they displace. This raises an intriguing additional possibility of harm: US science suffers because immigrants are less well connected to US science than the native born they displace. Essentially, scientist immigration could weaken the domestic knowledge networks that are critical to US scientific advancement.

To motivate our empirical approach to testing for such harm, we first develop a simple model of the market for scientists in which a sufficient condition of the absence of immigration induced harm to domestic science (as opposed to domestic scientists) is that immigrants are as well connected to domestic science as the native born they might displace. To test this condition, we conduct a hypothetical experiment in which each immigrant scientist is assumed to fully displace an appropriately matched US scientist. We then examine the impact on US science by comparing the subsequent citations by US scientists to the publications of the immigrant to the (hypothetically displaced) US scientist. In the model, the combination of displacement and differential spillovers could harm US science. However, a sufficient condition for the absence of harm is the absence of differential spillovers. We compare the relative citation patterns of domestic and foreign scientists. We find that immigrant scientists dramatically increase their number of US originating citations to their work upon moving to the US. Once in the US, immigrant and domestic scientists show no significant difference in US-based forward-citation connections. However, we do observe lower US-based spillovers from scientists that both move to universities with more co-nationals and that arrive from countries where english is uncommon.

We further refine our analysis to focus on scientists in the right tail of the productivity distribution, as a growing literature within the economics of science has demonstrated that these so-called "star scientists" may generate larger externalities than the median scientist (Azoulay et al., 2010; Waldinger, 2010; Oettl, 2012). With this subsample of elite scientists, the US-based spillovers deficit disappears.

We structure the remainder of the paper as follows. In the next section we develop a simple model of the market for scientific labor that provides a useful framework for the examination of the welfare implications of scientist immigration. The model allows for both domestic scientist displacement and differential spillovers from domestic and immigrant scientists. We describe our empirical strategy in Section 3 and our data and matching methodology in Section 4. Section 5 sets out our results. We conclude in Section 6 with a discussion of the limitations of our findings.

2 A Model of the Market for Scientists with Displacement and Differential Spillovers

We develop a simple model of the market for scientists in a given country and examine factors influencing the social welfare implications of immigration. The model allows for the displacement – or "crowding out" – of domestic scientists as a result of the immigration of scientists. We adopt the ex ante social welfare perspective of the receiving country, and thus ignore the welfare gains to immigrant scientists. Social welfare is thus measured by aggregate social surplus accruing to non-immigrant domestic residents; we do not focus on the distribution of that surplus. The model also allows for possible differential spillovers from domestic and immigrant scientists. We show it is possible for domestic social welfare to be harmed by immigration as a result of displacement if the difference between domestic and immigrant spillovers is large enough, even if immigration expands the overall size of the active scientific workforce. However, we show that a sufficient condition for immigration to improve domestic social welfare is that there is no difference in the size of per-scientist spillovers between domestic and immigrant scientists.

2.1 Basic market setup

We begin with specifications for labor supply and labor demand in the market for scientific labor. For simplicity, we assume that the units of labor are homogenous and each unit is a working scientist, although we later allow for differential spillovers between domestic and immigrant labor units.[1] The supply of domestic scientists, $L^s_{domestic}$, is a positive linear function of the wage, w:

$$L^s_{domestic} = \phi_0 + \phi_1 w. \tag{1}$$

Immigrant labor units, I, are supplied perfectly inelastically (possibly due to visa-related limitations), so the total supply of labor is[2]:

[1] The model is easily extended to allow for broader heterogeneity by defining labor units in efficiency (i.e., productivity-adjusted) units. Spillovers then also would be measured per efficiency unit, so that more productive scientists are assumed to generate more spillovers.

[2] In an efficiency-unit version of the model, the level of immigration is also measured in efficiency units.

$$L^s_{total} = \phi_0 + \phi_1 w + I. \tag{2}$$

Total labor demand, L^d, is a negative function of the wage:

$$L^d = \theta_0 - \theta_1 w. \tag{3}$$

The inverse of the labor demand function is also the marginal private value function. However, we also assume that there are positive spillovers associated with each unit of scientific labor employed. The per-scientist spillover (or externality) is equal to z (≥ 0), which is initially common across domestic and immigrant scientists. The marginal social value relationship is then given by:

$$MSV = \frac{1}{\theta_1}(\theta_0 - L) + z. \tag{4}$$

2.2 Baseline social surplus in the absence of immigration

As a preliminary step to establishing the effects of immigration on the market for scientific labor, we first examine the market equilibrium and social welfare in a no-immigration baseline. We graph the market equilibrium in Figure 1. The equilibrium wage and employment levels are:

$$w^* = \frac{\theta_0 - \phi_0}{\phi_1 + \theta_1}. \tag{5}$$

$$L^* = \frac{\phi_0\theta_1 + \phi_1\theta_0}{\phi_1 + \theta_1}. \tag{6}$$

Total social surplus from trade in the scientific labor market is the area between the inverse labor supply curve and marginal social value curve up to the equilibrium quantity of labor. This surplus is equal to:

$$\begin{aligned} S^* &= \int_0^{L^*} \left[\frac{1}{\theta_1}(\theta_0 - L) + z - \frac{1}{\phi_1}(L - \phi_0)\right] dL. \\ &= \left(\frac{\phi_0\theta_1 + \phi_1\theta_0}{\phi_1 + \theta_1}\right)\left[\left(\frac{\phi_0\theta_1 + \phi_1\theta_0}{2\phi_1\theta_1}\right) + z\right]. \end{aligned} \tag{7}$$

The total social surplus is given by the sum of areas A, B, and C in Figure 1. The existence of the positive externality means that the market equilibrium employment level is lower than the efficient (i.e., social-surplus-maximizing) level, where the latter is determined by the intersection between the labor supply curve and the marginal social value curve.

2.3 Social surplus with immigration but with identical spillovers for domestic and immigrant scientists

We next allow for positive immigration but initially assume that spillovers, z, are identical for domestic and immigrant scientists. We graph this case in Figure 2. The new equilibrium wage and employment levels are:

$$w^{**} = \frac{\theta_0 - \phi_0 - I}{\phi_1 + \theta_1}. \tag{8}$$

$$L^{**} = \frac{\phi_0\theta_1 + \phi_1\theta_0 + \phi_1 I}{\phi_1 + \theta_1}. \tag{9}$$

It is also useful to identify the employment level of domestic scientists at the new equilibrium with immigration:

$$L^{***} = \phi_0 + \phi_1 w^{**} = \frac{\phi_0\theta_1 + \phi_1\theta_0 - \phi_1 I}{\theta_1 + \phi_1}. \tag{10}$$

Notice that the domestic displacement is equal to:

$$L^{*} - L^{***} = \frac{\phi_1}{\phi_1 + \theta_1} I. \tag{11}$$

There is no displacement if ϕ_1 is equal to zero, so that the domestic labor supply is perfectly inelastic. To determine total social surplus, it is useful to separate out the surplus due to domestic versus immigrant scientists. Using Equation (10), the part due to domestic scientists is given by:

$$
\begin{aligned}
S^{**}_{domestic} &= \int_0^{L^{***}} \left[\frac{1}{\theta_1}(\theta_0 - L) + z - \frac{1}{\phi_1}(L - \phi_0)\right] dL \\
&= \left(\frac{\phi_0\theta_1 + \phi_1\theta_0 - \phi_1 I}{\phi_1 + \theta_1}\right)\left[\left(\frac{\phi_0\theta_1 + \phi_1\theta_0}{2\phi_1\theta_1}\right) + z + \frac{I}{2\theta_1}\right] \\
&= S^* - \left(\frac{\phi_1 z}{\phi_1 + \theta_1}\right) I - \left(\frac{\phi_1}{2\theta_1(\phi_1 + \theta_1)}\right) I^2,
\end{aligned}
\tag{12}
$$

where the last line makes use of Equation (7).

Because we are taking the perspective of the welfare of the receiving country, we exclude the surplus accruing directly to immigrant scientists. Domestic social surplus accruing from immigrants is thus the difference between the marginal social value curve and the post-immigration wage line (Equation (8)), where it is assumed that immigrants are the marginal labor suppliers. This surplus is given by:

$$
\begin{aligned}
S^{**}_{immigrant} &= \int_{L^{***}}^{L^{**}} \left[\frac{1}{\theta_1}(\theta_0 - L) + z - w^{**})\right] dL. \\
&= zI + \left(\frac{1}{2\theta_1}\right) I^2.
\end{aligned}
\tag{13}
$$

Total social surplus is found by summing the two components. After some cancellation, this yields:

$$
S^{**}_{total} = S^{**}_{domestic} + S^{**}_{immigrant} = S^* + \left(\frac{\theta_1 z}{\phi_1 + \theta_1}\right) I + \left(\frac{1}{2(\phi_1 + \theta_1)}\right) I^2. \tag{14}
$$

Noting that total social surplus depends positively on both the level and the square of the level of immigration, the surplus is increasing at an increasing rate with the level of immigration. The size of the gain will also depend positively on the size of the per-unit spillover, z, with a positive interaction between the size of the spillover and the level of immigration. The gain in social surplus is shown by the area enclosed by the dark black line in Figure 2.

2.4 Social surplus with immigration but with differential spillovers for domestic and immigrant scientists

We next examine the case where the spillover from domestic scientists, $z^D(\geq 0)$, differs from the spillover from immigrant scientists, $z^I(\geq 0)$, where it is assumed that $z^D \geq z^I$. The total social surplus is now:

$$S^{**}_{total} = S^{**}_{domestic} + S^{**}_{immigrant} = S^* + \left(\frac{\theta_1 z^I - \phi_1(z^D - z^I)}{\phi_1 + \theta_1}\right) I + \left(\frac{1}{2(\phi_1 + \theta_1)}\right) I^2. \tag{15}$$

Compared to the case of equal spillovers, an examination of Figure 3 shows a loss of social surplus on units that would have been supplied by domestic scientists in the absence of displacement. The lower spillovers from immigrant scientists also reduces the size of the gain from immigration, although there is still a direct gain in social surplus that is increasing non-linearly in the level of immigration. The overall impact on social surplus will depend on the relative sizes of these gains and losses. If the gap between z^D and z^I is large enough, it is possible that the displacement of domestic scientists reduces social surplus overall, notwithstanding the larger total size of the scientific workforce.

We now can identify from Equation (15) two distinct sufficient conditions for immigration not to reduce domestic social surplus given any level of immigration (i.e., for $S^{**}_{total} \geq S^*$). First, there will be no harm if there is no domestic displacement, i.e., $\phi_1 = 0$. Second, and central to the empirical part of the paper, there will be no harm if there is no difference between the domestic and immigrant spillover, i.e., $z^D - z^I = 0$.

Using Equation (15), we also can identify the necessary and sufficient condition for the absence of harm from immigration. This condition is:

$$z^I \geq \frac{\phi_1 z^D}{(\phi_1 + \theta_1)} - \left(\frac{1}{2(\phi_1 + \theta_1)}\right) I. \tag{16}$$

The "break-even" level of immigrant spillover is then the level of z^I at which Equation (16) holds with equality. We graph the break-even in Figure 4 as a function of the level of immigration.

The break-even level is declining in the level of immigration, reaching zero at an immigration level of $2\phi_1 z^D$. Given that the size of the immigrant spillover is assumed to be bounded from below at zero (i.e., the spillover is not negative), any immigration level above this level is associated with a net benefit regardless of the level of domestic displacement.

Summing up this section, we have found in the context of a simple market model with spillovers that it is possible that immigration harms domestic social welfare (as measured by the total surplus accruing to ex ante domestic residents from trade in the scientific labor market). This result requires both the displacement of domestic scientists by immigrants and lower spillovers from immigrants compared with domestic counterparts. However, the size of the spillover required from immigrant scientists to avoid immigration harming social welfare is decreasing in the level of immigration. Notwithstanding displacement effects, a sufficient condition for scientist immigration not to reduce ex ante domestic social welfare in the model is therefore an absence of differential spillovers.

As presented, the model applies to the general market for scientists. One could apply a narrower version to the segment of the market limited to employment at leading research universities. Displacement is then more naturally thought of as domestic scientists moving to lower-ranked universities, as found for example in Borjas and Doran (2012) as a result of the inflow of ex-Soviet mathematicians. In this case, we still would expect spillovers from displaced domestic scientists. However, if we assume that a faculty position in a leading university provides a privileged position in terms of the opportunities for relationship/network development[3] – and that domestic scientists are culturally or linguistically better positioned to take advantage of those opportunities – then downward institutional displacement could still be associated with a loss of aggregate spillovers and social welfare that again must be weighed against the direct gains from scientist immigration. The search for evidence on possible differential spillovers from domestic and immigrant scientists motivates the empirical work in the remainder of the paper.

[3]For example, positions at leading universities may provide faculty members with more graduate students. The pool of former graduate students then becomes a natural pool for matching with collaborators. In Agrawal et al. (2015), we develop a model in which scientists form the best match from the pool of former graduate students. Even where each potential former graduate student collaborator is drawn from a given uniform distribution, simply having more graduate students – and thus more draws – increases the expected value of collaboration. We then show that improvements in collaboration technology, which we assume to scale up the value of collaboration, are more valuable for scientists with more graduate students and thus more draws from which to find the best match.

3 Empirical Strategy

In the model of Section 3, a sufficient condition for the absence of harm is that foreign recruits are equally connected to US scientists compared to the domestic scientists they displace. This holds true even with full displacement. Our empirical strategy is to conduct a hypothetical experiment in which a foreign recruit displaces a matched domestic scientist, where the matching is done based on productivity, career age and discipline. We then compare the measures of local knowledge-flow supporting connections for the matched pairs. Our measure of connection is the subsequent forward citations by US scientists to the work of the immigrant/domestic match.

As a preliminary step, we first compare the pre-move connections of the eventual immigrants to their domestic matches. Looking just at the eventual immigrants, we then compare their connections to domestic scientists pre- and post move. Finally, the core of our analysis is a comparison of the post-move US connections of immigrants with their domestic matches. A finding of no difference would be consistent with the hypothesis of no harm to domestic science even with full displacement.

4 Data and Matching Methodology

Our primary objective is to compare spillover patterns between domestic versus immigrant scientists. Thus, we must identify scientists, their type (domestic or immigrant), and their spillovers. We use publication data to do this. Our primary source is the ISI Web of Science (WoS). We begin by collecting publications in six fields: 1) evolutionary biology, 2) mathematics, 3) economics, 4) neuroscience, 5) immunology, and 6) psychology. We collect all publications in the journals classified by the ISI Journal Citation Reports as being associated with each of those fields. In Table 1, we list the number of journals associated with each field and the number of papers we collect from this set of journals over the period 1979-2008. In terms of the number of publications, neuroscience and immunology are the two largest fields (825,048 and 639,439 papers, respectively) and evolutionary biology and psychology are the two smallest (114,190 and 191,333 papers, respectively). We present descriptive statistics of our star subsample in Table 2.

4.1 Identifying scientists

We conduct most of our analyses at the scientist-year level. So, using the publication data described above, we identify the set of scientists in each of the six fields. One data challenge with this process is that WoS data do not provide unique identifiers for scientists. In other words, the data do not distinguish between two different people who have the same name. Thus, we must disambiguate scientific authors. To do so, we employ an approach developed by Tang and Walsh (2010). The heuristic utilizes backward citations of focal papers to estimate the likelihood of the named author being a particular person. For example, if two papers reference a higher number of the same papers (weighted by how many times the paper has been cited, i.e., how popular or obscure it is), then the likelihood of those two papers belonging to the same author is higher. We attribute two papers to the same author if both papers cite two or more rare papers (fewer than 50 citations) in both papers. We repeat this process for all papers that list non-unique author names (i.e., same first initial and last name). We exclude scientists who do not have more than two publications linked to their name. In Table 1, we list the number of unique scientists we identify in each field. Once again, immunology and neuroscience are the two largest fields (84,649 and 91,405 scientists, respectively). The two smallest fields are evolutionary biology and psychology (9,619 and 9,805 scientists, respectively). Scientists enter the panel when they publish their first paper. We identify their location and status (star or not) on an annual basis.

4.2 Defining stars

We define stars as scientists in the 90th percentile in a given year and discipline in terms of their accumulated stock of citation-weighted paper output over the preceding years. To calculate a scientist's accumulated stock of citation-weighted paper output, we begin by identifying the set of papers they published in the years preceding the focal year. We then weight these papers by the number of citations they receive during our study period. For example, if a scientist published four papers by 1990 and these papers received 10, 20, 15, and 40 citations by 2008 (the final year of our study period), then that scientist's accumulated stock of citation-weighted paper output would be 85 in 1990. While we define a scientist's contribution on an annual basis, our measure or stardom is time-invariant whereby we classify a scientist as a star if the scientist has ever been above the 90th percentile (approximately 15% of scientists).[4] Furthermore, stars are defined relative to the other scientists in our sample in the same discipline. When we do analyses of the full sample (across all disciplines), we utilize the star categorization determined from the within-field analysis. Although citation practices vary across fields, scientists in the 90th percentile are disproportionately more productive than the median scientist across all fields as seen in Figure 5.

4.3 Identifying scientist locations

Using the unique author identifiers generated in the process described above for each paper, we then attribute each scientist to a particular institution for every year they are active. Scientists are active from the year they publish their first paper to the year they publish their last paper. Here again, we must overcome a data deficiency inherent within the WoS data; until recently, the WoS did not link institutions listed on an article to the authors. Instead, we impute author location using reprint information that provides a one-to-one mapping between the reprint author and the scientist's affiliation. In addition, we take advantage of single institution publications that allow us to directly link authors to institutions.

4.4 Defining immigration

With information on each scientist's location in each year, we identify the country of each scientist's institution. Domestic scientists are those who start their career in the U.S. and never emigrate. Immigrant scientists are those who start their career in a country other than the U.S. and some year after their first publication immigrate to the U.S.

4.5 Outcome measure

Our outcome measure if interest is knowledge flows. We identify all paper published by the focal scientist in the focal year for each scientist-year. From this set of papers, we count the number of

[4] Results are very similar if we conduct our analyses using a time-varying definition of star scientists whereby we only classify a scientist as a star in years in which their stock of citation-weighted paper output exceeds the 90th percentile.

forward citations (citations made to the focal paper by other papers in the future). We classify each forward citation as domestic if the first author of the future paper that references the focal paper is from the US and not-domestic otherwise.

4.6 Matching

Immigrant and domestic scientists may systematically differ along a range of dimensions hindering insightful comparisons between the two groups. As such, we identify a subset of both immigrant and domestic scientists that are on the common support of a vector of covariates related to scientific productivity in the year of the immigrant's move to the US. More specifically, for all immigrant scientists that immigrate to the US in year t we identify a domestic scientist match in year t that is in the same discipline, has a similar quality-weighted stock of publications, was equally as productive in year t, and has a similar career age. We make use of the of the Coarsened Exact Matching (CEM) methodology first developed by Iacus et al. (2012). Table 3 shows balance between immigrant and domestic scientists of our matched covariates across both the full and star sample.

5 Results

5.1 Comparisons of Matched Pairs

Our measure of connection is the number of forward citations by US scientists. Under the hypothetical scenario of full displacement of an equivalent domestic scientist, we test for significant differences between the subsequent forward citations to the work of the immigrant and their (hypothetically displaced) domestic matches. We look separately at all immigrants and also the subset of immigrant stars.

For each sample, it is informative to make three distinct comparisons. First, we compare pre-move immigrants with their domestic matches. This allows us understand the different degree of connectivity to US science before the move takes place. Second, we compare pre- and post-move immigrants. This allows us to understand the way the immigrant's connection to US science changes on moving to the US. And third, we compare post-move immigrants with their domestic matches. This is our main comparison, and it allows us understand how local knowledge spillovers would be affected even with full displacement of an equivalent domestic scientist.

Figure 6 provides a useful graphical depiction of all three comparisons and also allows for comparisons across the full and star samples. The general picture that emerges is that pre-move immigrants are significantly less connected to US scientists than their domestic matches as measured by forward citations to their work by US scientists. However, this gap tends to disappear with the move as immigrants appear to quite rapidly integrate into US science. Post move, forward citations to the immigrant's work are at least as large as their domestic matches.

Tables 4 and 5 provide formal tests for our three comparisons for the full and star samples respectively. The top panel of each table compares pre-move immigrants with their domestic matches. We look separately at total citations, US citations and the share of US citations in total citations. Indicating the success of the matching procedure, there is no significant difference in total citations for the immigrants and their domestic matches. However, the domestic matches have significantly higher US citations and higher shares of US citations in total citations. The difference in favor of the domestic matches in terms of US forward citations is particularly pronounced for the star sample, where on average the domestic matches have more than 10 additional US forward citations compared to the pre-move immigrants.

The bottom panel of each table tests for differences in citations to the work of immigrants pre- and post-move. Post-move immigrants have significantly more total citations and US citations and also have a higher share of US citations in their total citations for both the full and star samples.

The middle panel of each table compares post-move immigrants with their domestic matches – our central comparison. Post-move immigrants now have a larger number of US forward citations compared to their domestic matches (difference in full sample = 0.47, p-value = 0.15; difference in star sample = 2.58, p-value = 0.06). While the domestic matches still display a higher share of US citations in their total citations, the difference is not statistically significant.

Overall, using US forward citations as our measure of connection, immigrants are found to be at least as well connected to US science as the matched domestic scientists that they hypothetically displace. At least by this measure of connection, there is no evidence that scientist immigration would harm US science even with full displacement.

5.2 Factors Mediating the Integration of Immigrant Scientists into US Science

Recognizing that not all immigrant scientists will be equally well positioned to generate US-destined knowledge spillovers, we next explore how sensitive our main result is to plausible factors mediating the connection of immigrants to US science networks. Where a factor is plausibly linked to a weaker (stronger) relationship to US scientists, a finding of a smaller (larger) "immigrant premium" gives us greater confidence that the difference between the matched pairs provide a meaningful measure of different spillover potential between the immigrants and domestic scientists they (hypothetically) displace.

We examine two candidate mediating factors. The first is the prevalence of co-nationals at the destination institution. A higher prevalence of co-nationals is likely to be associated with more limited connections to US scientists (McPherson et al., 2001). Such differential integration is supported by findings that co-ethnicity supports knowledge flows (see, e.g., Agrawal et al., 2008), so that the close proximity of co-nationals could reduce the incentive for the immigrant to form connections with US scientists. The second is where the use of English is common in the immigrant's country of origin. Proficiency in English should be positively associated with the ability of the immigrant to connect with US scientists. A large literature has documented that proficiency in English is positively associated with success in English-speaking destination-country labour markets (see, e.g., Chiswick and Miller, 1995; Dustmann and Fabbri, 2003).

The results of these difference-in-difference analyses are shown in Tables 6 and 7. We focus in particular on the difference in post-move US cites between the immigrant and their domestic matches for both the full and star samples (Columns 2 and 5). In Table 6, we first compare the size of this "immigrant premium" where the immigrant has at most a single diaspora colleague with the case where they have two or more such colleagues. For the full sample, where immigrants are relatively isolated there is a statistically significant positive immigrant premium; but there is a statistically significant negative premium where the immigrant is co-located with two or more diaspora colleagues. The null of no difference between these premiums is strongly rejected (p-value

= 0.001). For the star sample, the size of the positive premium for the relatively isolated immigrants is even more pronounced than in the full sample. However, the effect is not statistically significant where there are two or more diaspora colleagues. The null of no difference between these premiums is again strongly rejected (p-value = 0.001).

In Table 7 we repeat these comparisons of the "immigrant premium" based on whether the immigrant comes from a country where the use of English is common or not. For the full sample, the premium is not statistically significant where the immigrant comes from a country where English is common. However, this premium is negative and statistically significant where the immigrant comes from a country where English is not common. The p-value for the null of no difference between the two cases is 0.072. Interestingly, for the star sample, we cannot reject the null of no difference between the two cases (p-value = 0.310). This may reflect the fact that strong English ability is common among stars regardless of whether they come from a country where the use of English is common or not.

Overall, the results of these difference-in-difference analyses are generally consistent with our priors. Immigrant scientists tend to perform better in terms of connections to US science when they are relatively isolated from co-nationals and also come from countries where the use of English is common – although the latter effect is not evident for stars.

6 Concluding Comments

The search for evidence of native wage and employment displacement effects has been a major theme of the immigration literature. More recently, in an attempt to better identify the benefits of high skilled immigration, more attention has focused on knowledge spillovers to native workers. But this raises a new possibility of harm if local knowledge networks are disrupted by arrivals that displace domestic workers that are better embedded in the knowledge sharing networks. To explore the possibility of such displacement, we use forward citation patterns in this paper to answer a simple question: Are immigrant scientists less connected to US scientists than the domestic scientists they displace? We find that although immigrant scientists are significantly less well connected to US scientists than their domestic matches pre immigration, the convergence to the level of connectivity observed for the domestic matches is rapid. Overall, we do not find evidence of harm to domestic science through a knowledge network disruption channel.

We conclude by noting some possible limitations of our findings and important areas for further research. First, while we use state-of-the-art matching techniques to identify our domestic matches for immigrant scientists, there is an inevitable residual concern that actual scientists displaced by immigrant arrivals are better connected to domestic scientists than these identified matches. In addition, it may also be that universities engaged in recruiting immigrant scientists are selecting those which are most likely to increase their productivity after arrival, increasing both the total knowledge spillovers they produce and also those that flow to the US.

Second, while we believe that forward citations provide the best measure of knowledge connections between scientists, other possibilities exist. One alternative is co-authorships with US scientists. Preliminary results suggest that immigrant scientists have fewer post-arrival co-authorship relationships with US scientists than their domestic matches. But conditional on a co-authorship relationship with a US scientist, the quality of the output as measured by forward citations to the work is higher for the immigrant-domestic collaborations. The nature of this quantity/quality tradeoff and also the relative importance of citation and co-authorship metrics as measures of connections between scientists requires further exploration.

Third, the diaspora and English language results point to the kind of variables that mediate the integration of immigrant scientists into US knowledge networks. More work is needed to better understand the integration process and the public or organizational policies that might facilitate it.

Finally, although scientists who publish are a key component of US knowledge networks, further work is required to confirm that immigration-related network disruption effects do not cause greater harm in other knowledge sectors. An advantage of examining scientific papers is that a natural paper trail of connections is provided through citations patterns. Patent citations may allow for an extension of the approach used here to explore network disruption effects in other parts of the of the US knowledge system.

References

Agrawal, Ajay, Devesh Kapur, and John McHale, "How do spatial and social proximity influence knowledge flows? Evidence from patent data," *Journal of Urban Economics*, 2008, *64* (2), 258 – 269.

_ , Iain Cockburn, and John McHale, "Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships," *Journal of Economics Geography*, 2006, *6* (5), 571–591.

_ , John McHale, and Alexander Oettl, "Collaboration, Stars, and the Changing Organization of Science: Evidence from Evolutionary Biology," in "The changing frontier: rethinking science and innovation policy," The University of Chicago Press, Chicago, October 2015. University of Toronto mimeo.

Azoulay, Pierre, Joshua Graff Zivin, and Jialan Wang, "Superstar Extinction," *Quarterly Journal of Economics*, 2010, *125* (2), 549–589.

Borjas, George, "The Labor-Market Impact of High-Skill Immigration," *The American Economic Review*, 2005, *95* (2), 56–60.

_ , "Do Foreign Students Crowd Out Native Students from Graduate Programs," in Ronald G. Ehrenberg and Paula E. Stephan, eds., *Science and the University*, University of Wisconsin Press, Madison, 2007.

_ and Kirk B Doran, "The Collapse of the Soviet Union and the Productivity of American Mathematicians*," *The Quarterly Journal of Economics*, 2012, *127* (3), 1143–1203.

Borjas, George J, *Immigration economics*, Harvard University Press, 2014.

Chiswick, Barry R and Paul W Miller, "The endogeneity between language and earnings: International analyses," *Journal of labor economics*, 1995, *13* (2), 246–288.

Dustmann, Christian and Francesca Fabbri, "Language proficiency and labour market performance of immigrants in the UK," *The Economic Journal*, 2003, *113* (489), 695–717.

Freeman, Richard, Ina Ganguli, and Raviv Murciano-Goroff, "Why and Wherefore of Increased Scientific Collaboration," in "The changing frontier: rethinking science and innovation policy," The University of Chicago Press, Chicago, 2015.

Hunt, Jennifer and Marjolaine Gauthier-Loiselle, "How Much Does Immigration Boost Innovation?," *American Economic Journal: Macroeconomics*, 2010, *2* (2), 31–56.

Iacus, Stefano M., Gary King, and Giuseppe Porro, "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 2012, *20* (1), 1–24.

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

Kerr, Sari Pekkala and William R Kerr, "Economic impacts of immigration: A survey," *Finnish Economic Papers, Finnish Economic Association 1-32*, Spring 2011, *24* (1), 1–32.

Kerr, William R and William F Lincoln, "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention," *Journal of Labor Economics*, 2010, *28* (3), 473–508.

Lan, Flora, Katherine Hale, and Emilda Rivers, "Immigrants' Growing Presence in the U.S. Science and Engineering Workforce: Education and Employment Characteristics in 2013," NSF 15-328, National Science Foundation 2015.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, 2001, *27* (1), 415–444.

National Academies of Sciences, Engineering, and Medicine, *The Economic and Fiscal Consequences of Immigration*, The National Academies Press, Washington, DC, 2016.

Oettl, Alexander, "Reconceptualizing stars: Scientist helpfulness and peer performance," *Management Science*, 2012, *58* (6), 1122–1140.

Peri, Giovanni, "The effect of immigration on productivity: Evidence from US states," *Review of Economics and Statistics*, 2012, *94* (1), 348–358.

_ , Kevin Shih, and Chad Sparber, "STEM workers, H1B Visas and Productivity in US Cities," 2013. Working Paper.

Stephan, Paula, *How Economics Shapes Science*, Cambridge: Harvard University Press, 2012.

Tang, Li and John P Walsh, "Bibliometric fingerprints: Name disambiguation based on approximate structure equivalence of cognitive maps," *Scientometrics*, 2010, *84* (3), 763–784.

Thompson, Peter and Melanie Fox-Kean, "Patent citations and the geography of knowledge spillovers: A reassessment," *American Economic Review*, 2005, *95* (1), 450–460.

Waldinger, Fabian, "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany," *Journal of Political Economy*, 2010, *118* (4), 787–831.

Figure 1: Market Equilibrium and Total Social Surplus, No Immigration

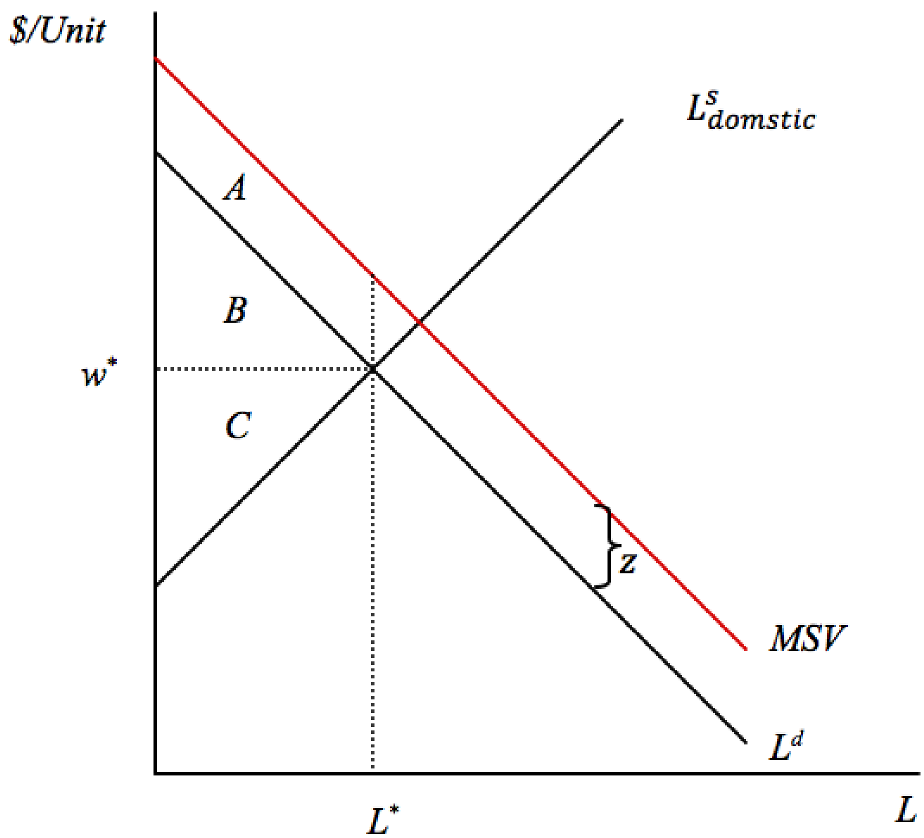
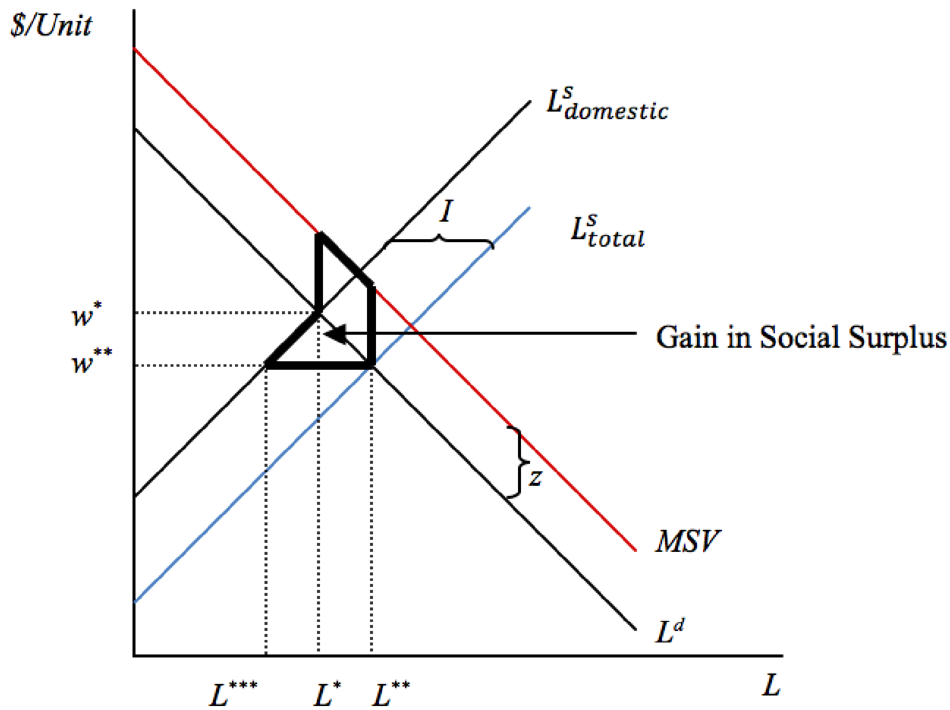


Figure 2: Market Equilibrium and the Gain in Social Surplus from Immigration



Note: The per-scientist externality is assumed to be equal to z for domestic and immigrant scientists.

Figure 3: Market Equilibrium and the Gain and Loss of Social Surplus when the per Scientist Externality is Lower for Immigrant Scientists

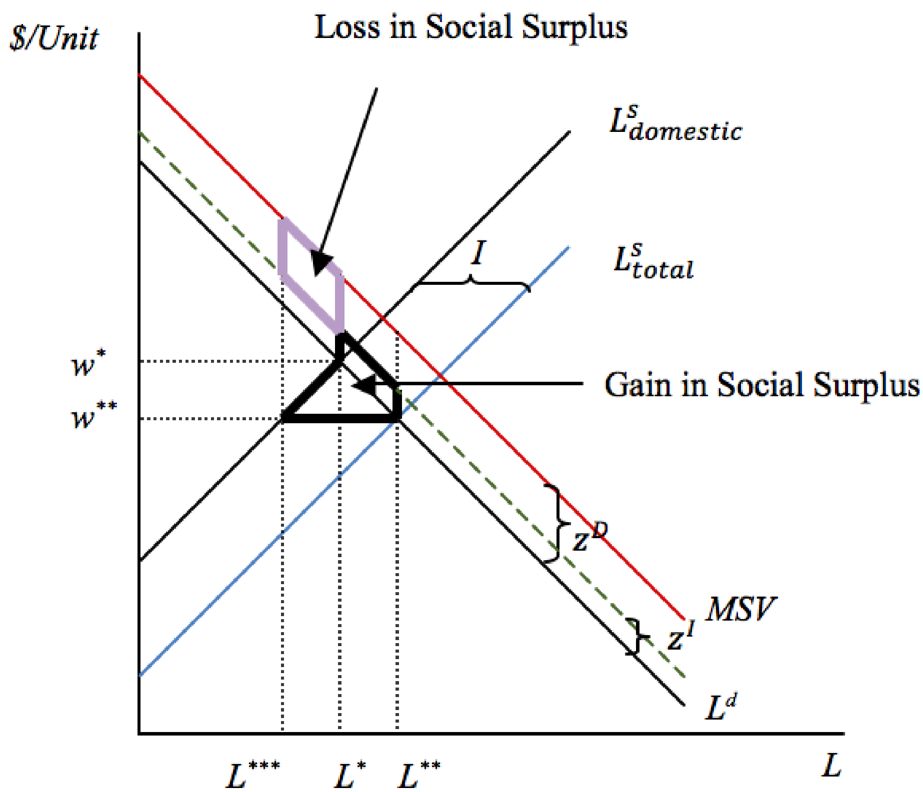


Figure 4: The Level of the Per-Scientist Externality for Immigrant Scientists for No Change in Social Surplus to Occur as a Result of Immigration

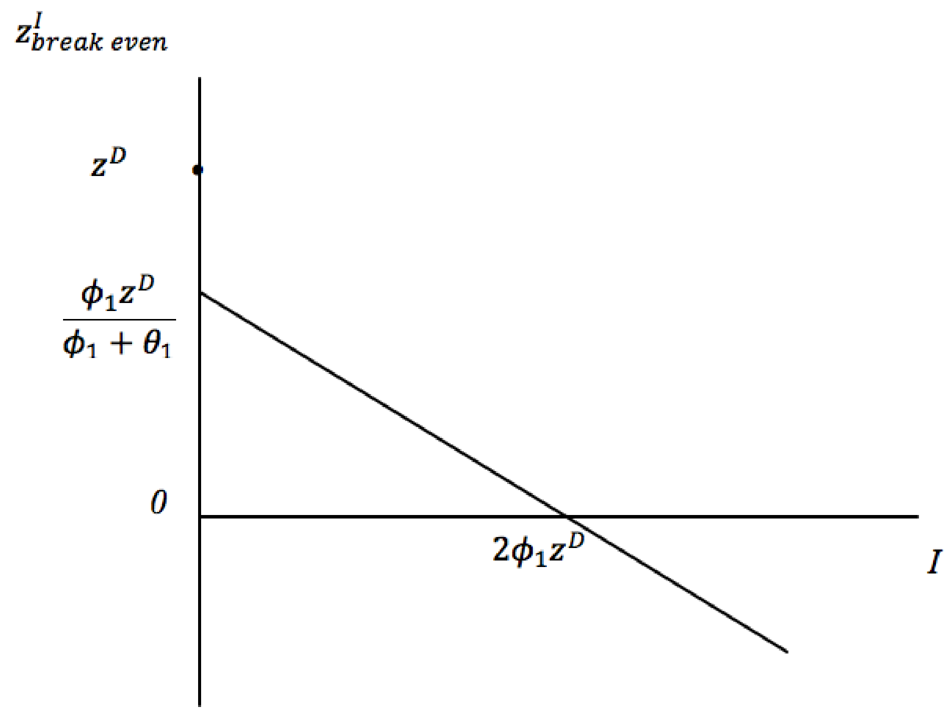


Figure 5: Citation Stock Percentiles by Field in 1995

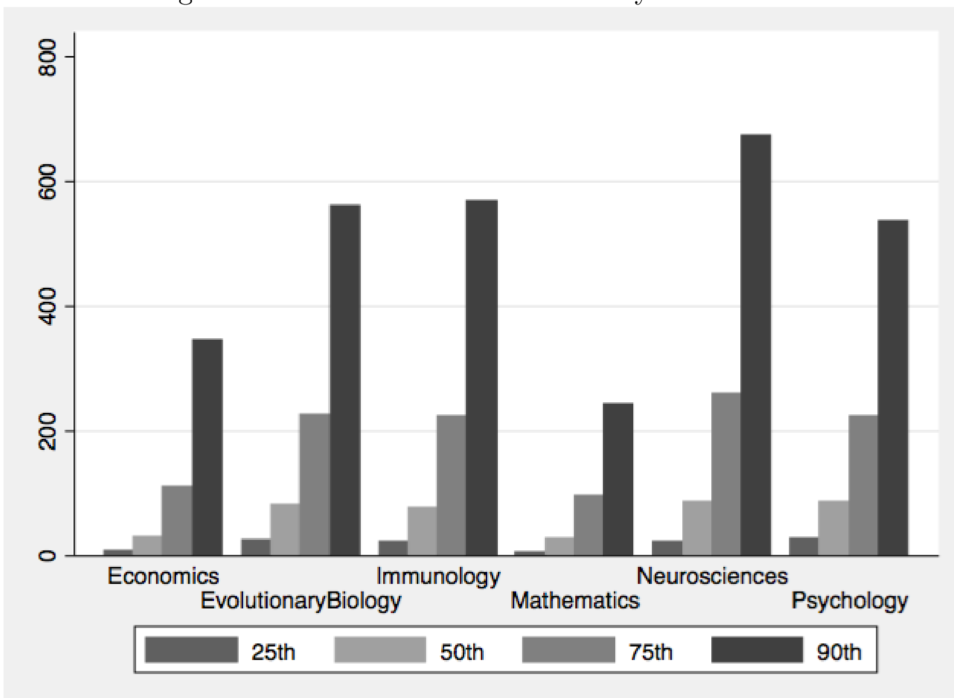
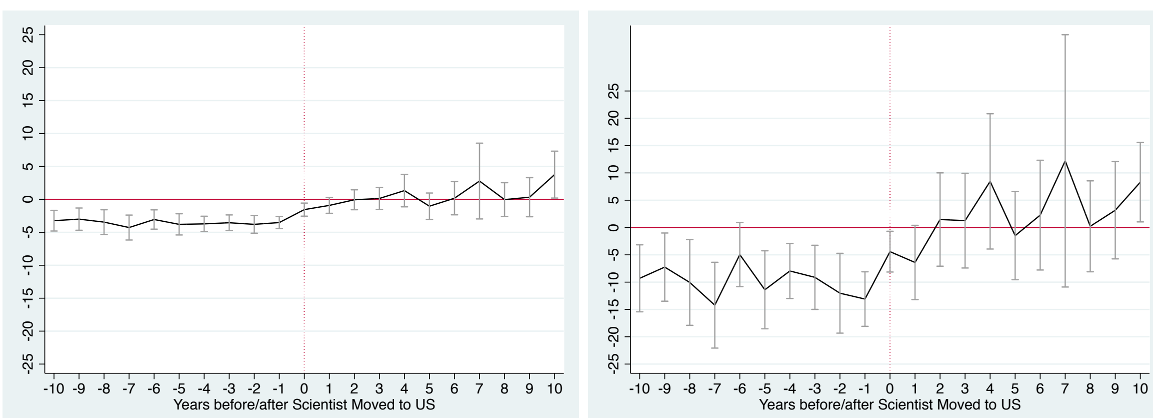


Figure 6: Number of US Citations: Immigrants relative to Domestics



(a) Full Sample

(b) Star Sample

Notes: This figure plots point estimates for leading and lagging indicators for the migration of a scientist to the US. Both panels plot the point estimates of the following specification estimated using OLS: $USCitations_{it} = \sum_{\tau=0}^{10} \alpha_{-\tau} Arrival_{i,t-\tau} + \sum_{\tau=1}^{10} \alpha_{+\tau} Arrival_{i,t+\tau} + \sum_{\tau=0}^{10} \beta_{-\tau} Arrival_{i,t-\tau} \times immigrant_i + \sum_{\tau=1}^{10} \beta_{+\tau} Arrival_{i,t+\tau} \times immigrant_i + \theta(Age_{it}) + \delta_t + \varepsilon_{it}$. *US Citations*$_{it}$ is the number of citations received by scientist i in year t from US-authored papers. The α parameters (21 in all) controls for the US citation patterns of the matched domestic scientists for each year 10 years prior and post to the matched immigrants arrival. The β parameters are our point estimates of interest and are the ones plotted in the above figure. These reflect the differences in US citation patterns between immigrants and domestic scientists for each year around the move year (+/- 10 years). θ flexible controls for scientist i's age and δ is a full set of year dummies. There is no constant in this specification. The vertical bars correspond to 95% confidence intervals with scientist-clustered standard errors.

Table 1: Descriptive Statistics

Discipline	Journals	Papers	Scientists	Domestics	Immigrants	Citations/ Scientist/ Year	Coauthors/ Scientist/ Year
Economics	214	105,305	18,466	10,302	552	8.38	0.39
Evol. Biology	42	55,035	9,619	4,497	286	18.76	0.74
Immunology	175	586,424	84,649	35,281	3,311	16.17	2.59
Mathematics	190	126,535	22,156	7,644	1,065	3.67	0.42
Neuroscience	247	678,572	91,405	38,074	4,209	19.14	2.14
Psychology	71	49,316	9,805	5,495	218	6.9	0.67
Total	939	1,601,187	236,100	101,293	9,641	12.17†	1.16†

Notes: *Scientists* refers to the total number of scientists active in the world. *Domestics* refers to the number of US-based scientists that started their careers in the US. *Immigrants* refers to the number of US-based scientists that emigrated to the US. Note that *Domestics* and *Immigrants* do not sum to *Scientists* as we do not report counts of scientists in the rest of the world that do not emigrate to the US during our study period. The last two columns count the mean number of citations received / unique coauthors per scientist per year.
† Means, instead of sums, are reported for these two columns.

Table 2: Descriptive Statistics (Star Sample)

Discipline	Journals	Papers	Scientists	Domestics	Immigrants	Citations/ Scientist/ Year	Coauthors/ Scientist/ Year
Economics	214	29,727	1,324	1,058	101	34.45	0.72
Evol. Biology	42	14,866	755	458	49	59.72	1.21
Immunology	175	131,385	7,220	4,094	687	53.71	4.71
Mathematics	190	39,369	1,653	893	214	12.06	0.76
Neuroscience	247	144,420	7,129	3,902	799	61.72	3.83
Psychology	71	16,530	801	548	46	20.58	1.00
Total	939	376,297	18,882	10,953	1,896	49.69†	3.34†

Notes: *Scientists* refers to the total number of scientists active in the world. *Domestics* refers to the number of US-based scientists that started their careers in the US. *Immigrants* refers to the number of US-based scientists that emigrated to the US. Note that *Domestics* and *Immigrants* do not sum to *Scientists* as we do not report counts of scientists in the rest of the world that do not emigrate to the US during our study period. The last two columns count the mean number of citations received / unique coauthors per scientist per year.
† Means, instead of sums, are reported for these two columns.

Table 3: Descriptive Statistics of Domestic and Immigrant Scientists

Variable	Domestic Scientists mean	Immigrant Scientists mean	difference	p-value of difference
Panel A				
Career Age	7.43	7.48	-0.05	0.73
Ever a Star	0.14	0.14	0	1
$\sum^{t-1}$ Cites	154.11	153.92	0.19	0.97
Cites	34.36	35.19	-0.83	0.37
Observations	4,623	4,623		
Panel B: Star Sample	N=640	N=640		
Career Age	9.99	10.1	-0.11	0.79
Ever a Star	1	1	0	1
$\sum^{t-1}$ Cites	449.9	442.67	7.23	0.78
Cites	78.7	80.83	-2.13	0.54
Observations	640	640		

Table 4: Mean Comparisons of Citations

	Variable	Immigrant Mean (1)	Immigrant Std. Dev. (2)	Domestic Mean (3)	Domestic Std.Dev (4)	Column Diff (5)	p-value of diff (6)
	Pre-Move Period	*N = 28,449*		*N = 28,449*			
(1)	Citations	17.68	54.25	17.92	49.64	-0.24	0.58
(2)	US Citations	6.55	24.58	10.13	28.30	-3.57	0.00
(3)	Share of US Citation	0.33	0.22	0.57	0.23	-0.24	0.00
	Post-Move Period.	*N = 21,008*		*N = 21,008*			
(4)	Citations	20.45	74.40	18.08	52.19	2.36	0.00
(5)	US Citations	10.25	39.06	9.77	28.26	0.47	0.15
(6)	Share of US Citations	0.49	0.25	0.54	0.24	-0.06	0.00
		Row Diff	p-value of diff				
(7)	Citations	2.77	0.00				
(8)	US Citations	3.70	0.00				
(9)	Share of US Citations	0.15	0.00				

Notes: Each observation is at the scientist-year level. Citations is the mean sum of the number of forward citations to papers published by the scientist in the specific time period (pre or post move). US Citations is the mean annual count of the number of forward citations to papers published by scientist i in the time period where the first author of the citing paper resides in the US. Immigrant and Domestic scientists are matched using coarsened exact matching along the following dimensions: scientist age, total citations within the US, and discipline.

Table 5: Mean Comparisons of Citations (Star Sample)

	Variable	Immigrant Mean (1)	Immigrant Std. Dev. (2)	Domestic Mean (3)	Domestic Std.Dev (4)	Column Diff (5)	p-value of diff (6)
	Pre-Move Period	*N = 5,103*		*N = 5,103*			
(1)	Citations	52.28	108.92	53.08	96.50	-0.79	0.70
(2)	US Citations	19.95	50.90	30.17	55.18	-10.21	0.00
(3)	Share of US Citations	0.35	0.18	0.58	0.18	-0.23	0.00
	Post-Move Period	*N = 4,611*		*N = 4,611*			
(4)	Citations	59.04	144.97	50.01	95.22	9.03	0.00
(5)	US Citations	29.79	76.53	27.22	51.49	2.58	0.06
(6)	Share of US Citations	0.49	0.20	0.55	0.20	-0.06	0.00
		Row Diff	p-value of diff				
(7)	Citations	6.75	0.00				
(8)	US Citations	9.84	0.00				
(9)	Share of US Citations	0.15	0.00				

Notes: Each observation is at the scientist-year level. Citations is the mean sum of the number of forward citations to papers published by the scientist in the specific time period (pre or post move). US Citations is the mean annual count of the number of forward citations to papers published by scientist i in the time period where the first author of the citing paper resides in the US. Immigrant and Domestic scientists are matched using coarsened exact matching along the following dimensions: scientist age, total citations within the US, and discipline.

Table 6: Difference-in-Differences, Diaspora Effect

	(1)	(2)	(3)	(4)	(5)	(6)
Sample		Full			Star	
Dependent Variable	Cites	US Cites	US Cites Share	Cites	US Cites	US Cites Share
Immigrant with ≤ 1 diaspora colleagues†	0.255**	0.187*	−0.131**	0.447**	0.377**	−0.123**
	(0.079)	(0.084)	(0.027)	(0.127)	(0.137)	(0.038)
Immigrant with ≥ 2 diaspora colleagues‡	−0.086	−0.165**	−0.198**	−0.001	−0.07	−0.183**
	(0.056)	(0.058)	(0.026)	(0.088)	(0.085)	(0.039)
University Fixed Effects	✓	✓	✓	✓	✓	✓
p-value of $H_0 : \dagger = \ddagger$	0.000	0.001	0.031	0.000	0.001	0.188
Observations	41722	41664	41664	9161	9161	9161

Notes: The unit of analysis is the scientist-year. The sample consists of domestic and immigrant scientists in the US. All specifications are estimated using poisson quasi maximum likelihood. Robust standard errors clustered at the scientist level are in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 7: Difference-in-Differences, Countries where English is Common

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Full			Star		
Dependent Variable	Cites	US Cites	US Cites Share	Cites	US Cites	US Cites Share
Immigrant from country where english is common†	0.120^{+}	0.049	-0.121^{**}	0.088	0.019	-0.167^{**}
	(0.068)	(0.068)	(0.028)	(0.095)	(0.092)	(0.039)
Immigrant from country where english is uncommon‡	-0.048	-0.128^{+}	-0.232^{**}	0.242^{+}	0.173	-0.159^{**}
	(0.063)	(0.068)	(0.026)	(0.129)	(0.141)	(0.040)
University Fixed Effects	✓	✓	✓	✓	✓	✓
p-value of $H_0 : \dagger = \ddagger$	0.065	0.072	0.000	0.261	0.310	0.850
Observations	41722	41664	41664	9161	9161	9161

Notes: The unit of analysis is the scientist-year. The sample consists of domestic and immigrant scientists in the US. All specifications are estimated using poisson quasi maximum likelihood. Robust standard errors clustered at the scientist level are in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$