The Ethnic Migrant Inventor Effect:

Codification and Recombination of Knowledge Across Borders

Prithwiraj Choudhury and Do Yoon Kim[1]

Ethnic migrant inventors may differ from locals in terms of the knowledge they bring to host firms. We study the role of first-generation ethnic migrant inventors in cross-border transfer of knowledge previously locked within the cultural context of their home regions. Using a unique dataset of Chinese and Indian herbal patents filed in the United States, we find that an increase in the supply of first-generation ethnic migrant inventors increases the rate of codification of herbal knowledge at U.S. assignees by 4.5 percent. Our identification comes from an exogenous shock to the quota of H1B visas and from a list of entities exempted from the shock. We also find that ethnic migrant inventors are more likely to engage in reuse of their prior knowledge, whereas knowledge recombination is more likely to be pursued by teams comprising inventors from other ethnic backgrounds.

Keywords: skilled migration; ethnic migration; first-generation migrant; cultural context; knowledge flows; knowledge reuse; knowledge recombination; recombinant creation; H1B visas

[1] We thank Rajshree Agarwal, Gautam Ahuja, Ashish Arora, Iain Cockburn, Alfonso Gambardella, Shane Greenstein, Elena Kulchina, Mariko Sakakibara, Minyuan Zhao, and seminar participants and discussants at Harvard Business School, London Business School's Sumantra Ghoshal Conference, the NBER Productivity Seminar, The Wharton School and the University of Michigan for comments on a prior draft. Corresponding author: Prithwiraj (Raj) Choudhury, pchoudhury@hbs.edu

We live in an age of global mass migration on the part of skilled knowledge workers. Recent literature on skilled migration has pointed out that the number of U.S. migrants with tertiary degrees rose by nearly 130 percent between 1990 and 2010 (Kerr *et al.*, 2016). However, this phenomenon coexists with significant concern about migrants displacing jobs for the native-born, and about whether countries should reduce their intake of skilled migrants. This paper sidesteps the debate on whether skilled migrants create or pre-empt employment opportunities for locals. Instead, we study an interesting phenomenon related to the role of ethnic migrant inventors in innovation at their host firms. We argue that ethnic migrant inventors differ from locals and play an important role in transferring to the host firm, knowledge previously *locked within* the cultural context of their home regions. We study recombination of such knowledge once it has been transferred to the host firm.

To motivate this line of research, we present a stylized example involving the role of Dr. Hari P. Cohli, an ethnic skilled migrant and a researcher in immunology at the University of Mississippi, in transferring knowledge about Indian herbal medicine to the west. Dr. Cohli migrated from India to enroll at the University of Toronto and later studied and worked at SUNY Buffalo and the Johnson Space Center in Houston. At the University of Mississippi, Dr. Cohly met Dr. S.K. Das, a plastic surgeon who was about to amputate the leg of a patient with a wound that would not heal due to a condition known as restenosis, a gap between two blood vessels. In his native city of Agra, Dr. Cohly had attended the Indian herbal medicinal (*Ayurveda*) discourses of Dr. M.B. Lal Sahab, a religious teacher and Edinburgh-educated parasitologist. Drawing on these discourses, Dr. Cohly suggested using turmeric to heal the wound. When the patient recovered, avoiding amputation, Dr. Cohly and Dr. Das conducted a clinical trial and filed a U.S. patent for "a wound-healing agent consisting of an effective amount of turmeric powder." This example illustrates the role of an ethnic migrant researcher in codifying knowledge of Indian medicinal herbs and transferring it to the west.

Economic history is replete with such examples of skilled ethnic migrants' transfers of knowledge previously locked within the cultural context of their home regions to the host region and to their host organizations. Scholars have documented the case of Russian mathematicians' transfers of knowledge about such fields as partial differential equations and symplectic topology to the United States, after the fall of the Soviet Union in the 1990s (Borjas and Doran, 2012; Ganguli, 2015). Another example is the seventeenth-century migration of Huguenots from France to Brandenburg-Prussia. As Hornung (2014) states, the Huguenots introduced a great variety of advanced skills and new technologies to their host region. One account lists 46 professions introduced by Huguenots, mostly textile-related, all of which were previously unknown to the host country. One Huguenot carried with him the secret of dyeing fabrics in a special way; another brought the art of printing on cotton. Economic history has also documented the role of the Venetian city council in luring skilled migrant artisans from Florence by offering them patents, which protected their monopoly rights to the knowledge they possessed (Belfanti 2004). After the expiration of the patent, the Florentine artisans joined the local artisan guild in Venice and shared their knowledge with local artisans.

Despite these and other examples (discussed in a later section) of skilled ethnic migrants transferring to the host region knowledge previously locked within the cultural context of their home regions, the recent literature on high-skilled immigration has focused on the employment effects of skilled immigration, namely whether skilled migrants create or displace jobs for the native-born (Kerr, Kerr, and Lincoln, 2015; Kerr and Lincoln, 2010; Doran, Gelber, and Isen, 2016). This question is part of a larger debate in the strategy-and-innovation literature on the global sourcing of talent. Scholars such as Lewin, Massini, and Peeters (2009) have long argued that the shortage of highly skilled technical talent in the United States, and the need to access qualified knowledge workers abroad, should drive western firms' offshoring innovation decisions. This literature tends to

take for granted that migrant inventors *resemble* (and thus can replace) local inventors; the literature thus overlooks the possibility that migrant inventors *differ from* local inventors. Specifically, migrant inventors may differ from local inventors in terms of the knowledge they transfer across borders—knowledge that could subsequently be recombined by other inventors, including local inventors.

The field of innovation and strategy has produced a rich literature on the role of context in shaping innovative outcomes (Hambrick and MacMillan, 1985). We draw on this literature, and the literature on cross-national variation in context along cultural, linguistic, religious, and other dimensions (Ghemawat, 2001; Berry, Guillén, and Zhou, 2010), to posit that knowledge can be deeply embedded in its cultural, linguistic, or religious context. We also draw on the literature on impediments to the transfer of sticky knowledge (Von Hippel, 1994; Szulanski, 1996) to argue that, ex-ante, such knowledge can be locked in its home region and thus unavailable to knowledge production in an ethnic migrant inventor's host country.

Given this argument, we study the role of ethnic migrant inventors, especially first-generation ethnic migrants, in transferring such knowledge from their home regions to their host firms. We use the term *home regions* to denote the regions where the knowledge in question was first discovered and/or put into use. We use the phrase *ethnic migrants* solely to denote residents of those regions who migrate to the host firm. Such individuals could be first-generation ethnic migrants (i.e., individuals who migrate themselves) or later-generation ethnic migrants (whose forebears migrated). For purposes of this paper, the phrase *non-ethnic inventors* refers to individuals who did not migrate to the host firm from the home regions of the ethnic migrant inventors. They include both local inventors and those who migrated to the host firm from other regions of the world.

We argue that first-generation ethnic migrant inventors have both the absorptive capacity (Cohen and Levinthal, 1990) and the career incentives (Holmström, 1999) to transfer to their host firms, knowledge that was previously locked within the cultural contexts of their home regions. We

also draw on literature on the mechanisms of knowledge recombination (Allen 1977, Fleming 2001) to study knowledge recombination once such knowledge has been transferred to the host firm. We argue that ethnic migrants, whose career incentives stress quick wins (Amabile and Kramer, 2011), tend to reuse prior knowledge, while teams whose members include non-ethnic inventors respond to incentives for knowledge recombination.

To test these propositions, we created a unique dataset of 758 granted herbal patents filed with the U.S. Patent Office (USPTO) between 1977 and 2013 by U.S. firms and universities that were operating during a 1999–2003 immigration shock. The filing entities included such large western multinationals as Abbott Laboratories, Amgen, Eli Lilly, Pfizer, Colgate Palmolive, Dow Chemical, Proctor & Gamble, and Unilever, as well as large U.S. universities. Access to high skilled migrants is important in this industry, as evidenced by responses to the tightening of U.S. immigration policies: for instance, Kenneth Frazier, the CEO of Merck, asserted that "We have to get the best scientists, the best employees around the world" (Johnson, 2017).

Our identification strategy is driven by an exogenous shock to U.S. H1B employment visas. In 1998 and 2000, Congress passed legislation which temporarily increased the quota of H1B visas. The new legislation also exempted universities and a select list of other entities from the quota. Exploiting this policy change to estimate a difference-in-differences model, we find that an increase in the supply of first-generation ethnic migrant inventors increases the rate of codification of herbal knowledge at U.S. assignees. We observe a 4.5 percent increase in herbal patenting at firms subject to the visa cap.[2] Furthermore, inventors of Chinese and Indian ethnicities are more likely to engage in knowledge reuse; meanwhile, teams comprising non-ethnic inventors are more apt to engage in knowledge recombination (i.e., combining herbs with other synthetic compounds to create relatively

[2] It is plausible that demand side effects lead to an increase in herbal patenting over time; however, given that the visa shock only affects capped firms, our analysis is focused on estimating the effect of a supply shock on herbal patenting.

novel formulations). To our knowledge, we are the first scholars to exploit the H1B exclusion list for purposes of research.

Our findings contribute to the literatures on skilled migration and ethnic migration (Kerr, 2008; Foley and Kerr, 2013; Franzoni, Scellato, and Stephan, 2014) by suggesting that ethnic migrant inventors differ from local inventors with regard to the knowledge they bring to a firm, knowledge that can subsequently be recombined. This insight should be incorporated into policy debates on skilled migration. Our findings also inform the strategy-and-innovation literature on inventor mobility and knowledge flows (Breschi and Lissoni, 2009; Oettl and Agrawal, 2008; Rosenkopf and Almeida, 2003; Song, Almeida, and Wu, 2003) and on the microfoundations of knowledge recombination within firms (Carnabuci and Operti, 2013; Fleming, 2001; Gruber, Harhoff, and Hoisl, 2013).

Theory and Hypotheses

Ethnic migrant inventors and transfer of knowledge across borders

The strategy and international business literature has documented cross-national variation in contexts along cultural, linguistic, religious and other dimensions (Ghemawat 2001, Berry *et al.* 2010). We build on this literature to argue that knowledge can be *locked* within a given cultural context, and, as a corollary, that such knowledge may not transfer easily to regions whose cultural contexts are dissimilar.

To make this argument, we draw on the literature on impediments to transferring sticky knowledge (Von Hippel, 1994; Szulanski, 1996; Szulanski, 2002 Jensen and Szulanski, 2004). Szulanski (2002) outlined several impediments to knowledge transfer having to do with the characteristics of the knowledge, including that it might be ex-ante unproven in a new context and/or that causal ambiguity might prevail about the effectiveness of employing it. Knowledge might be considered unproven in certain locales because it is codified in a different language. In the

case of Chinese and Indian herbal medicine, western researchers are apt to lack access to knowledge codified in Mandarin and Sanskrit textbooks; they might also be skeptical about using such knowledge, given the paucity of clinical trials proving the effectiveness of herbs. Such knowledge might also suffer from *causal ambiguity*, Szulanski's (2002) term for lack of knowledge about why a given set of actions results in a given outcome ("know-why"). In the case of Chinese and Indian herbal medicine, uncertainty might be concentrated around whether Chinese/Indian herbs would be effective in the western climatic environment. Uncertainty of this kind could prevent effective knowledge transfer. We also leverage the theory of tacit knowledge (Polanyi, 1966, Dasgupta and David, 1994, Cowan and Foray, 1997) to argue that the "know-how" involved in using knowledge embedded in the cultural context of a given country may not transfer easily across borders. Specifically, tacit know-how might be embedded in the actions of expert practitioners and thus available only to immediate observers. Such experts might all be concentrated in the home region, where the knowledge is locked.

Ethnic inventors who migrate to a new region could be instrumental in transferring such knowledge across borders. We argue that ethnic migrant inventors have the absorptive capacity to gainfully employ knowledge previously locked within the cultural context of their home regions. Building on Cohen and Levinthal (1990), we argue that, through prolonged prior exposure, ethnic migrant inventors are likely to have acquired deep understanding of both the know-why and the know-how aspects of knowledge locked in the cultural context of their home regions. They may also have had access to experts who possessed pertinent tacit know-how. In addition to possessing the absorptive capacity to transfer knowledge across borders, ethnic migrant inventors have incentives to codify such knowledge at their host firms. We build on theory about the career incentives of individuals within firms (Holmström, 1999) to argue that transferring and codifying knowledge from their home regions enables ethnic migrants to establish reputations as high performers at their host

firms.[3] It is important to note that both the absorptive capacity to transfer knowledge across borders and the incentives to codify such knowledge at the host firm are apt to be especially marked among *first-generation* ethnic migrant inventors. Compared to their second-generation (or longer-settled) ethnic migrant counterparts, first-generation ethnic migrants have had more home-country exposure to pertinent knowledge and to experts who harbor such knowledge and have stronger incentives to distinguish themselves at their new host firms. In other words, when first-generation ethnic migrant inventors move across borders, we can expect to see an increase in the codification of knowledge previously locked within the cultural context of their home countries.[4] This leads us to our first hypothesis:

Hypothesis 1: An increase in the supply of first-generation ethnic migrant inventors increases the rate of codification at their host firms of knowledge previously locked within the cultural context of their home region.

Recombination of knowledge

Next, we consider how knowledge previously locked within the cultural context of a home region gets *recombined* after its transfer to the host firm, as well as the roles of ethnic migrant inventors and non-ethnic inventors in knowledge reuse and recombination.

The study of knowledge recombination has a rich tradition in the fields of economics and strategy (Schumpeter, 1939; Nelson and Winter, 1982; Henderson and Clark, 1990). One stream of this literature focuses on the microfoundations of knowledge recombination, i.e., the role that individuals play in knowledge recombination within a given firm. This tradition dates back to Allen (1977) and is framed by Fleming (2001) as the process of recombinant search that is characteristic of individual inventors. In the subsequent literature, Carnabuci and Operti (2013) designate two distinct

[3] Holmström (1999) outlines how an individual's concern for a future career may influence his or her incentives to put in effort or make decisions on the job. In the model, the person's productive abilities are revealed over time through observations of performance, and an implicit contract links today's performance to future wages.

[4] In addition to directly codifying such knowledge, first-generation ethnic migrants might transfer it to ethnic and non-ethnic peers at their host firms, who might over time codify variants of it. This insight builds on prior research tracing how mobile inventors might act as cross-cultural brokers, imparting knowledge to co-located peers after geographic mobility (Burt 1992; Oettl and Agrawal, 2008; Almeida and Kogut, 1999; Singh 2005).

recombinant search strategies: "recombinant creation" (creating recombinations new to the firm) and "recombinant reuse" (reconfiguring combinations already known to the firm). In our context recombination is more likely to consist of recombinant creation, or recombining knowledge components from the western context with knowledge components previously embedded in the ethnic migrant inventor's home region. We build on this literature to theorize that, in the case of knowledge transferred from the cultural context of a home region to a host firm, ethnic migrant inventors are more likely to engage in knowledge reuse than in either form of recombination; by contrast, teams that include non-ethnic inventors are more likely to engage in recombinant creation.

Building on the prior literature about absorptive capacity and career incentives, we argue that the incentives driving ethnic migrant inventors tend to promote reuse of knowledge over recombination. In our setting, we frame knowledge reuse as the ethnic migrant inventor's appropriation of knowledge from the home country and codification of that knowledge in the new context of a western firm. Given her relatively high absorptive capacity with respect to knowledge previously locked in her home country, the ethnic migrant can realistically expect a high likelihood of success at reusing home-country knowledge. We build on expectancy theory (Vroom, 1964) to argue that this expectation is likely to motivate reuse of such knowledge by ethnic migrants. Also, ethnic migrant inventors (especially first-generation ethnic migrants) at host firms experience strong incentives to establish a reputation quickly. The literature on career incentives has shown that *quick wins* help boost individuals' confidence and instill a sense of career progress (Amabile and Kramer, 2011; Connelly *et al.*, 2011). Arguably, transferring and reusing home-country knowledge could be a quick win for ethnic migrants. In contrast, recombining the same knowledge with components well known to the western firm would require the ethnic migrant inventor to develop the absorptive capacity to assimilate knowledge components from *both* settings. We also know from prior literature that developing the absorptive capacity to work with new knowledge components calls for repeated

exposure to related problems over the course of many practice trials. Hence, recombination is unlikely to result in a *quick win* for ethnic migrants.

We also theorize that non-ethnic inventors might be more inclined to engage in recombinant creation by working with ethnic migrant inventors and/or by working on their own. As prior literature has shown, to engage in recombinant creation, inventors need to have knowledge diversity (Cohen and Levinthal, 1990; Ahuja and Morris Lampert, 2001; Carnabuci and Operti, 2013).[5] When non-ethnic inventors collaborate with ethnic migrants, the resulting team would have greater knowledge diversity than either of the inventor groups working separately. Recombinant creation could also emerge from non-ethnic inventors working alone. Non-ethnic inventors might learn the new knowledge being codified from their ethnic peers and might recombine such knowledge. This insight builds on prior research tracing how mobile inventors might act as cross-cultural brokers, imparting knowledge to co-located peers after geographic mobility (Burt 1992; Oettl and Agrawal, 2008; Almeida and Kogut, 1999; Singh 2005). However, working on a recombinant creation project is a risky choice for an individual. As Holmström (1999) states, the risk preferences that govern choices of projects are driven by individuals' career concerns. Non-ethnic inventors might make the risky choice of working on recombinant creation, given that information about their talent is more likely to be known to the firm. In short, we expect higher reuse to emerge from teams composed exclusively of ethnic migrants. By contrast, we expect recombinant creation to emerge from teams whose members include non-ethnic inventors. This leads to our second hypothesis:

[5] According to Cohen and Levinthal (1990), knowledge diversity facilitates the innovative processes of individual inventors by promoting novel associations and linkages pertinent to the problems they are attempting to solve. In a similar vein, Ahuja and Morris Lampert (2001) have shown that knowledge diversity helps individuals find radically novel approaches to solving technological problems. Extending this argument, Carnabuci and Operti (2013) theorize that knowledge diversity helps individual inventors engage in recombinant creation.

Hypothesis 2: Teams composed solely of ethnic migrants are more likely to reuse knowledge from their home regions at their host firms; knowledge recombination is more likely to be pursued by teams that include non-ethnic inventors.

Data, Variables, and Identification Strategy

To test our hypotheses, we use a unique dataset of Chinese and Indian herbal patents filed in the United States. For several reasons, herbal patents are an appropriate empirical setting in which to study the transfer and recombination of knowledge previously locked within the cultural context of ethnic migrant inventors' home regions. For centuries, both China and India have accumulated extensive knowledge about the medicinal properties of herbs, within medical canons distinct from the western medical canon (*Ayurveda, Unani, Siddha, Yoga*, and TCM, or Traditional Chinese Medicine). Chinese and Indian migrant knowledge workers account for more beneficiaries of temporary U.S. work visas than any other national groups. In the pharmaceutical industry, which generates many herbal patents, immigrants represent 33% of the total R&D workforce and 43% of medical/life scientists (Michel and Witte, 2014). This scenario represents an opportunity to test whether knowledge of Chinese and Indian herbal medicine is transferred to the west by first-generation migrant Chinese and Indian scientists, and whether non-ethnic inventors play an important role in recombination of that knowledge.

A unique dataset of herbal patents

Within the entire universe of USPTO patents, we identified herbal patents filed between 1977 and 2013. We categorized a patent as herbal if the application named at least one herb and specified its use. Our search process consisted of three iterative steps. First, we first obtained a list of 52 herbs, and their common and scientific names, from the National Center for Complementary and Alternative Medicine (NCCAM) website. We then searched Thompson Innovation and LexisNexis TotalPatents for USPTO patents whose abstract or title named any of these herbs. We found 7,163

such patents. We then iteratively searched the identified patents for more herb names and collected additional herbal patents. The addition of herb names extracted from those patents ultimately produced a list of 1,785 herbs. The total number of patent-herb pairs exceeds the number of herbal patents because a single patent can name multiple herbs. The most frequently named herbs appear in Table A17 in the Appendix. Next, we performed a classification search using both the International Patent Classification (IPC) and the U.S. Patent Classification (USPC) schemes.[6] Finally, we used the Traditional Chinese Medicine (TCM) database to augment our dataset and read patent abstracts to further validate our list and identify additional herb names and their uses; we appended all patents with U.S. priority to our existing dataset.

We then collected information about the patent assignees from USPTO's PatentsView and Capital IQ. Our identification strategy, detailed below, focused on U.S. entities that *could have* hired inventors between 1999 and 2003; thus, we concentrated on patents filed by U.S. assignees (firms and universities) that could have hired inventors in 1999–2003. To determine which assignees fell into this category, we used PatentsView's assignee classification data and data from Capital IQ on the locations of the assignees' headquarters. Our sample consists of 1,794 patent filings filed by U.S.-based assignees, 981 of which were granted. Capital IQ also provided firms' founding dates, allowing us to impute a firm's age.[7] We dropped 536 patents whose assignees (mostly firms) were founded after the visa shock (and thus not affected by it) or that stopped patenting before 1999, the first year of the shock, suggesting that the assignee may not have been in operation during the visa-

[6] In particular, we used the IPC class A61K36+ and the USPC classifications 424/725 and 514/783. The IPC class was introduced in 2002 by a Committee of Experts at IPC Union for purposes of linking the Traditional Knowledge Research Classification (TKRC) with IPC as part of the work of the World Intellectual Property Organization Traditional Knowledge (WIPO-TK) Task Force.

[7] For assignees without an equivalent firm in Capital IQ, we used the earliest patent application year in the USPTO database as the assignee's founding year. Our results are robust to other definitions of assignee founding years, and to including the assignee in the sample, even prior to the first patent filing date, with lags.

shock period. Doing so left 1,258 patent filings.[8] In the base case, we kept only patent applications that were granted, resulting in 758 patents submitted by 401 assignees.

Finally, we aggregated our herbal patent data at the assignee-year level by counting the number of herbal patents granted to a given assignee in each year. We used the *tsfill* command in Stata to fill in missing assignee-year-level observations. For years when an assignee was not granted any herbal patents, we set the assignee-year observation at zero.[9] We dropped any assignee-year pairs that corresponded to dates before an assignee's founding date. The assignee-year level dataset is thus an unbalanced panel consisting of 8,998 observations (an average of 22.4 years of observations for 401 assignees).

Identification strategy and variables

We proposed two hypotheses: that an exogenous increase in the supply of first-generation ethnic migrant inventors increases the rate of codification of knowledge previously locked within their home countries; and that while ethnic migrant inventors are more likely to reuse knowledge transferred across borders, recombination is more likely to be pursued by teams comprising non-ethnic inventors. To test the first hypothesis, we run a difference-in-differences model on log herbal patent counts at the assignee level to determine whether herbal patenting changes as the supply of Chinese/Indian migrants to the U.S. changes. To test the second hypothesis, we test whether the probability of recombination correlates with the presence of Chinese/Indian inventors' names on a

[8] This number includes patents filed by entities that did not file patents during the visa shock but did so after it ended. Our results are robust to using the sample of assignees that filed patents during the visa-shock period. Section 7.5 in the appendix contains more details on the selection of assignees.

[9] The outcome variable, number of herbal patents granted to an assignee in a given year, has many zeros; and the variance of this variable is larger than its mean. Consequently, we investigated whether a Poisson regression is appropriate for this data. Following Cameron and Trivedi (2010), we checked whether the outcome variable looks like a Poisson-distributed random variable. We see a slightly higher probability of zeros in the observed outcome than the Poisson distribution would predict (0.0089), but no difference in predicted counts of zeros using Stata's countfit command. The contribution of zeros to the Pearson Chi-Square statistic is 0.001, further showing that our data does not suffer from over inflation of zeros. We also ran robustness checks using a Poisson and Poisson QML specification.

patent. The next section describes our natural experiment, defines variables, and presents empirical specifications.

A Natural Experiment: The H1B Visa Shock and Excluded Entities

The key barrier to identifying whether or not a supply shock of first-generation ethnic migrants leads to greater codification of knowledge is the existence of unobservables affecting both codification and immigration patterns. In our setting, the returns to investment in herbal patenting increased in the mid-1990s as the market for herbal remedies grew. This increase in demand may have led firms to accelerate herbal patenting and to recruit more experts on herbal remedies. Also, even if we found a correlation between ethnic migrant inventors and codification of knowledge, we would be unable to determine whether these ethnic inventors were first-generation migrants. The goal was to find an exogenous increase in the inflow of ethnic Indian and Chinese inventors to the United States, unrelated to determinants of herbal patenting.

In pursuit of this goal, we utilized an exogenous shock to skilled immigration to the United States. In 1998 and 2000, Congress promulgated two laws that differentially impacted some firms' capacity to recruit skilled labor from abroad. As a result, the number of H1B employment visas increased from 65,000 in 1998 to 115,000 in 1999 and 195,000 in 2001, and then dropped back to 65,000 in 2004. Both laws were responses to increased demand for information technology (IT) professionals during the dot-com bubble. Thus, the flow of first-generation migrants is plausibly exogenous to the filing of herbal patents since most of the workers filled IT-related positions. We focus on Chinese and Indian inventors because they are the two largest groups to receive H1B visas: workers from India account for the majority of H1B recipients, followed by workers from China. Figure 1 illustrates the cap on H1B visas over time. In the appendix (section 2.1), we verify that this H1B shock was meaningful for our sample of assignees by comparing the trend in Labor Condition Applications (LCAs)—a prerequisite filed by an employer that intends to apply for H1B visas—

submitted by our sample of assignees to the number of *new* unique ethnic names on the list of inventors in our sample during the shock period.

An interesting feature of this immigration shock, to our knowledge not previously exploited in academic research, is that certain entities were exempted from the visa cap: workers "(1) at an institution of higher education or a related or affiliated nonprofit entity, or (2) at a nonprofit research organization or a governmental research organization"[10] could hire as many employees as they wished via the H1B visa. In addition to universities, the cap-exempt assignee list includes firms like Amgen, Monsanto, and Pioneer.[11] Of our 401 assignees, 73 were exempt from the visa cap; they were granted 123 patents. This exemption allows us to study the differential effect of the visa-cap increase by comparing herbal patent grants at capped and exempt patenting assignees.

Dependent Variables

Patent counts

Our main dependent variable (*Log Patent Count*) is the log number of herbal patents granted to an assignee (a firm or university) in a given year. We aggregate our herbal patents into assignee-year-level observations. For years in which an assignee was included in the sample but was not granted any herbal patents, we set the number of patents at zero. Since the number of herbal patent grants in an assignee-year is skewed, we add 1 and take logs.[12]

Recombined

To create the dependent variable for Hypothesis 2, we code herbal patents as *Recombined* if the patent text refers to synthetic non-herbal formulations in addition to herbs. We use the Derwent

[10] Source: https://www.uscis.gov/sites/default/files/USCIS/Laws/Memoranda/Static_Files_Memoranda/Archives%201998-2008/2006/ac21c060606.pdf

[11] Firms were able to hire migrants in the 'cap-exempt' mode because the statute exempted from the numerical limitations of the cap those migrants who were employed "at" a qualifying institution, a broader category than employment "by" a qualifying institution. In other words, firms could also claim cap exemption in hiring a migrant because the migrant would perform duties "at" a qualifying institution.

[12] Our results are robust to using nonlinear count models, including negative binomial, Poisson and Poisson QML. As discussed earlier, our data does not suffer from over inflation of zeros.

classification, a manually curated standardized system for classifying patents maintained by Thomson Reuters. We identify three classes within the Derwent classification that consist of both herbs and synthetic compounds, and code a patent as recombined if it belongs to any of these classes. Section 7.1 in the Appendix explains how this variable was coded.

New inventors

PatentsView allows us to identify individual inventors in our dataset and to track them across time. We combine this data with our ethnicity classification, described below, to count the number of new Chinese/Indian inventors associated with each patent (*New Ethnic Inventors*).

Independent Variables

Inventor ethnicity

Though patent documents do not specify inventors' ethnicities, we were able to predict likely ethnicities using names' linguistic cues. Probabilistically, surnames like Xing are more likely to be associated with Chinese individuals than with other ethnicities. Building on this insight, we utilized the open-source name categorizer "ethnicityguesser" to determine inventors' ethnicities.[13] The software, based on the Natural Language ToolKit (NLTK) package in Python, comes prepackaged with a set of names and associated ethnicities. As a robustness check, we compared our ethnicity-classification results when using different training sets to Ambekar *et al.* (2009), who use state-of-the-art hidden Markov models and decision trees for classification. The Appendix reports correlations across our measures and other established measures of ethnicity classification (see Tables A12 and A13). We find correlations above 0.9 for all our ethnicity measures. We create indicator variables to specify whether the set of inventors on a patent is uniformly Chinese/Indian (*Fully Ethnic*), uniformly non-Chinese/Indian (*Non-Ethnic*), or of mixed ethnicity (*Mixed Team*).

Empirical Specifications

[13] GitHub kitofans/ethnicityguesser - https://github.com/kitofans/ethnicityguesser

To recap, Hypothesis 1 states that an increase in the supply of first-generation ethnic migrant inventors increases the rate of codification of knowledge previously locked within the cultural context of migrant inventors' home regions. Our identification comes from a difference-in-differences model using capped firms as the treated group; the treatment period is the visa-shock period. We estimate this regression equation using our assignee-year level data:

$$\boldsymbol{ln(1 + Patent\ Count_{jt}) = \alpha + \beta_1 Capped_j + \beta_2 Shock_t + \gamma Capped_j \times Shock_t + \phi_j + \lambda_t + \varepsilon_{jt}} \quad (1)$$

Our dependent variable $ln(1+Patent\ Count_{jt})$ is the log number of herbal patents filed by assignee j in year t. The variables $Capped_j$, $Shock_t$ are dummies for whether assignee j is subject to the H1B cap and whether the patent application year is within the treatment period, 1999–2003. To control for assignee-level unobservables and temporal unobservables, we include assignee fixed effects (ϕ_j) and year fixed effects (λ_t) in some specifications, in which case the variables $Capped_j$, $Shock_t$ are respectively dropped. Here, the β coefficients capture the time-invariant difference in patenting between capped firms and exempt firms (β_1), and the percentage change in herbal patenting over time (β_2). Our main coefficient of interest, the interaction term γ, captures the percent increase in herbal patenting caused by relaxing the H1B visa cap.

To test Hypothesis 2, which asserts that teams composed solely of ethnic migrants are more likely to reuse knowledge than to recombine it, we must document the relationship between knowledge recombination and inventor ethnicities. We test whether there is a significant association between a patent with only Chinese/Indian inventors and the nature of the knowledge created (i.e., recombination or reuse). We run the following regression equation using our patent-level data:[14]

[14] We also utilize the supply shock and estimate alternate specifications: $Recombined_{ijt}=\delta Capped_j \times Shock_t + \lambda_t + \phi_j + \varepsilon_{ijt}$. and $Recombined_{ijt} = \gamma Capped_j \times Shock_t \times FullyEthnic_i + \beta_1 Capped_j \times Shock_t + \beta_2 Shock_t \times FullyEthnic_i + \beta_3 Capped_j \times FullyEthnic_i + \alpha FullyEthnic_i + \lambda_t + \phi_j + \varepsilon_{ijt}$. We include year fixed effects (λ_t) and assignee fixed effects (φ_j). δ measures the change in likelihood of recombination caused by an increase in the flow of first-generation ethnic migrants, and γ measures the differential impact of the shock on patents with solely ethnic inventors.

$$\boldsymbol{Recombined_{ijt} = \beta Fully\ Ethnic_{ijt} + \phi_j + \lambda_t + \varepsilon_{ijt}} \quad (2)$$

Again, the dependent variable ($Recombined_{ijt}$) is coded as 1 if patent *i* by assignee *j* with application year *t* contains synthetic herbal compounds. $Fully\ Ethnic_{ijt}$ signifies that a patent's inventors all have Chinese/Indian names. We include assignee fixed effects (ϕ_j) to control for assignee-level unobservables that determine the likelihood that a patent mentions synthetic compounds. The β coefficient measures the association between ethnic composition and the probability of knowledge recombination. Because Hypothesis 2 states that patents with exclusively ethnic inventors are more likely to involve reuse, we should see $\beta < 0$ even after controlling for assignee fixed effects.

We are also interested in temporal trends in recombination. Specifically, we wish to learn whether, over time, knowledge is more likely to be recombined, and by which ethnic groups. In the base case, we estimate the following equation:

$$\boldsymbol{Recombined_{iht} = \alpha + \beta_1 Time\ Since\ Herb\ Introduced_{iht} +}$$
$$\boldsymbol{\beta_2 (Time\ Since\ Herb\ Introduced_{iht})^2 + \nu_h + \varepsilon_{iht}} \quad (3)$$

Our dependent variable is an indicator denoting whether patent i filed at time t uses herb h, and whether it is a recombination of herbs and synthetic compounds. Our independent variable measures how much time has passed since herb h was first introduced in the United States, measured using the time between the filing date of the first U.S. patent containing the herb and the filing date of patent *i*. We include herb fixed effects ν_h because we are interested in the within-herb effect of time. The β_1 coefficient with herb fixed effects thus measures, for a particular herb, whether it is more likely to be recombined initially ($\beta_1 < 0$) or after it has become more widespread ($\beta_1 > 0$). The β_2 coefficient estimates non-linear effects.

Results

Summary statistics for capped and exempt assignees

Table 1 compares the assignee-year-level characteristics of capped and exempt assignees, during and outside of periods characterized by a visa-cap increase. We report the means and standard deviations, and the results of a *t*-test with standard errors. Figures 2a and 2b document the increase in herbal patenting by U.S. entities during our study period: we see a significantly greater increase in herbal patenting between 1999 and 2003, when the visa cap was increased, at capped firms than at exempt firms. In Figure 2c, we observe a similar trend in the numbers of new Chinese/Indian inventors' names on patents. We will verify these trends below using robust econometric methods.

Testing Hypothesis 1: Difference-in-differences estimation

Hypothesis 1 states that an increase in the supply of first-generation ethnic migrant inventors increases the rate of codification of knowledge previously locked within the cultural context of those inventors' home regions. Table 2 presents the results of estimating our main difference-in-differences specification, equation (1) using OLS. Our dependent variable is the log of 1 plus the number of patent grants at the assignee-year level. Standard errors are clustered at the assignee level.

We see from the coefficient on the interaction term (*Capped x Shock*) in Column 1 that the visa shock increased the log herbal patent count ($\beta = 0.044$, $p = 0.010$). This suggests a 4.5 percent increase in herbal patenting at firms subject to the visa cap. The positive coefficient on the $Capped_j$ variable denotes that capped firms file more patents than exempt firms, ($\beta = 0.010$ $p = 0.026$), invariant to time. Similarly, we see from the positive coefficient on $Shock_j$ that the visa-shock is associated with an increase in herbal patenting, ($\beta = 0.048$ $p = 0.001$). Column 2 controls for assignee-year-level factors that may affect herbal patenting, such as assignee age and the number of Chinese/Indian inventors employed by the assignee in a given year. The coefficient on our main dependent variable is robust to controlling for year fixed effects (Column 3) and for assignee fixed effects (Column 4). For the last specification, we see that the visa shock caused herbal patenting to increase by 4.5 percent, with β=0.044, p=0.013. Our results are robust to relaxing the founding-year

assumptions discussed in the Appendix (Table A7) and to using nonlinear count models instead of OLS, which we report in the Appendix (Table A8). Separately, in Appendix table A18 we test whether stricter immigration policies after 2004 decreased herbal patenting and find a negative effect (β= -0.036, p<0.001), i.e. a 3.7 percent decrease in herbal patent filing at capped firms after reduction of the visa cap in 2004.

Testing for parallel trends and dynamics of the visa shock

To learn how the visa shock affected the two groups of assignees each year, and to see whether the common-trends assumption holds, we include interactions between capped-firm dummies and lead and lag terms for the implied visa shock. Graphically, we can plot how being in the capped sample affected herbal patenting by Chinese/Indian inventors over time by plotting the coefficients on all interaction terms, as in Autor (2003). Doing so also allows us to compare the treatment and control groups during the pre-treatment period. In Figure 3, we see that herbal patenting increased significantly between 1999 and 2003, when the visa cap was increased. Furthermore, with the exception of 1993 and 1994, all the coefficients on the interactions are statistically insignificant in the pre-shock period, supporting the difference-in-difference assumption of parallel trends.[15]

New-herb introduction

We have shown that a supply shock of first-generation migrants increases codification of herbal patents in the United States. Yet we do not know whether the knowledge being codified was unfamiliar to western assignees as suggested by H1, which hypothesized the codification of knowledge previously locked within the home region of ethnic migrant inventors. We estimate an OLS model for the probability that new inventors in the sample, especially new ethnic inventors, introduce a *new* herb and/or an herb unfamiliar in the west (as measured by Google Ngrams). The

[15] The positive and significant effects of 1993 and 1994 might be related to the Immigration Act of 1990 (enacted on November 29, 1990), which provided for 140,000 visas per year for job-based immigration in five categories (EB1, EB2, EB3, EB4 and EB5). It also created new categories of nonimmigrant visas (the O and P categories) for extraordinarily skilled foreigners in the realm of science.

results appear in Table A4 in the Appendix.[16] We find that new inventors are 12 percentage points more likely than existing inventors to introduce a new herb (p= 0.006); the presence of new ethnic inventors on a patent adds 15 percentage points to the likelihood of introducing a new herb (p= 0.041). Similarly, the participation of a new inventor is associated with approximately a one-standard-deviation decrease in the measure of prior familiarity of the herb in the west (p= 0.004); the participation of a new ethnic inventor is associated with an additional one-standard-deviation decrease (p= 0.047) in prior familiarity. Section 3 in the Appendix provides more detail.

Testing Hypothesis 2: Knowledge reuse and knowledge recombination

Hypothesis 2 states that ethnic migrant inventors are likely to reuse knowledge transferred from their home regions; knowledge recombination is more likely to be pursued by teams that include non-ethnic inventors. Table 3 presents the results of estimating Equation 2 using OLS and clustered standard errors, clustered at the assignee level. Column 1 compares patents with only Chinese/Indian names to patents with at least one non-ethnic inventor. We see that fully ethnic teams are 34.6 percentage points less likely to engage in recombination than teams that include non-ethnic inventors (p=0.026). Furthermore, we utilize the identification strategy as an exogenous supply shock in the likelihood of observing fully ethnic teams consisting exclusively of first-generation ethnic migrants. We estimate a difference-in-differences specification similar to equation (1) and interact it with an indicator for fully ethnic teams to derive the differential impact of the visa shock on the fully ethnic groups. Column 2 shows how the visa shock affected recombination; column 3 shows the differential impact of the visa shock on patents with only Chinese/Indian names (fully ethnic). Column 4 compares recombination probabilities for mixed teams and non-ethnic teams. All specifications include assignee and application year fixed effects.

Column 2 shows that inventors at firms that were subject to the visa cap are 31.3 percentage

[16] Specifically, we test $1(HerbIntro)_{ijt} = \beta_0 + \beta_1 NewInventor_i + \beta_2 NewEthnicInventor_i + \phi_j + \lambda_t + \gamma X_{jt} + \varepsilon_{ijt}$, and $NewHerbFamiliarity_i = \beta_0 + \beta_1 NewInventor_i + \beta_2 NewEthnicInventor_i + \phi_j + \lambda_t + \gamma X_{jt} + \varepsilon_{ijt}$

points less likely to recombine knowledge (δ = -0.313, p=0.022).[17] Furthermore, Column 3 shows a larger negative effect for fully ethnic teams at capped firms during the shock (γ=-0.897, p<0.001), suggesting that reuse is being driven by fully ethnic teams consisting of first-generation migrants. Finally, Column 4 shows that both mixed and fully non-ethnic teams are more likely to recombine than are fully ethnic teams (p=0.053 and p=0.021 respectively). A t-test fails to reject the null of no difference between these two coefficients. In each specification, we include various controls. We include the number of claims and the citation count to control for patent characteristics. Since recombination may be driven by the broader collaboration network of prior co-inventions, we control for average number of co-inventors, i.e. average centrality, and for prior exposure to ethnic co-inventors computed at the patent level (results available with authors upon request).

Recombination over time

Figure 4 (a) plots the likelihood of recombination over time after an herb is introduced to the United States (measured by the filing date of the first U.S. patent containing the herb), by inventor ethnicity, averaged across all herbs mentioned in the patent text.[18] The solid line plots the probability of recombination by any ethnicity, and is the sum of the dashed lines, which plot the probabilities of recombination by specific ethnicities. We see a slightly positive, nonlinear relationship for recombination in general, especially for non-ethnic inventors. Mixed teams seem to increase their rate of recombination later in an herb's life.

Next, we formally test for the relationship between recombination and the duration of time since an herb's introduction. In the base case, each observation is an herb-patent pair, for all herbs and for all patents. For example, a patent with three herbs would have three herb-level observations. Table 4 presents the results of estimating Equation 3. In all specifications, we include herb and year

[17] Our results are robust to nonlinear specifications using logistic regression. Results are available upon request.
[18] We use the *lpoly* command in Stata to plot the smoothed values of a Kernel-weighted local polynomial regression of the recombination probability on the length of time since the oldest herb was released.

fixed effects in addition to patent-level controls.[19] Column 1 shows a positive and significant relationship between recombination and time elapsed since an herb was introduced (i.e., the difference between the year of application for the focal patent and the first year an herb was used in any patent), with $\beta = 0.0053$ and $p<0.001$. Column 2 further shows an inverted-U-shaped relationship between recombination and time since an herb was introduced ($\beta_1 = 0.0188$, $p<0.001$, $\beta_2 =-0.0004$, $p<0.0001$). We also report estimates for recombination by non-ethnic inventors (Columns 3-4), mixed teams (5-6), and ethnic inventors (7-8). We observe an inverted-U-shaped relationship for recombination by non-ethnic inventors. For mixed teams, we see instead an increasing rate of recombination over time. We formally test for inverted U-shaped relationships as per Simonsohn (2017) and present results in Figure 4 (b-c). Intuitively, this method tests whether one observes both a positive and significant slope to the left of a given cutoff and a negative and significant slope to the right, with the cutoff being set via a Robin-Hood algorithm (which the author shows is robust to errors). For non-ethnic teams, recombination increases in the first 16.72 years (β_{low}=0.027, $p<0.001$) but decreases afterwards (β_{high}= -0.008, $p<0.001$). For fully ethnic teams, we observe significantly lower levels of recombination; an increase in the first 20.74 years (β_{low}=0.002, $p<0.001$) and decreases afterwards (β_{high}=-0.003, $p<0.001$). Mixed teams increase recombination over time and recombination seems to be disproportionately driven by teams with non-ethnic inventors and mixed teams.

Robustness checks and secondary analyses

The relevance of the shock to the sample of assignees. An empirical concern is that, if H1B visas were primarily reserved for IT companies, firms that applied for herbal patents might not have been affected by the visa shock. We performed several tests related to determine the relevance of the shock to our sample of firms. Since we cannot observe the number of H1B visa grants at the firm level, we use as a proxy Labor Condition Applications (LCAs, a prerequisite for H-1B visas), which

[19] We include citation counts, number of claims, number of inventors, assignee age at filing, and whether the patent has new inventors.

we observe for the universe of firms and thus for each firm in our sample. We match all U.S.-based assignees that were granted herbal patents to company names in LCA filings. Our control group is therefore all non-pharmaceutical companies that filed for LCAs. The control group includes companies such as IBM, Merrill Lynch and Goldman Sachs; herbal patent assignees include firms such as Amgen and Eli Lilly. Figure A2 in the Appendix plots the quantile-quantile plot of total LCAs filed by our assignees and by the control group. The quantile-quantile plot shows that the number of LCAs filed by herbal-patent assignees is left-skewed, suggesting that herbal assignees probably hired more people via H1B visas than did all other firms that filed LCAs. *t*-test results show that herbal assignees filed for 146.7 more LCAs on average (*t*-statistic 4.81) than other firms in the LCA sample, further showing that herbal-patent assignees were indeed a major beneficiary of H1B visas. We also performed back-of-the-envelope calculations of the effect of the visa shock on hiring to show that the H1B shock had a meaningful impact on inventor hiring and on patenting at the treated assignees (cap-subject assignees with herbal patent grants) in our sample. Our calculations suggest that hiring ten additional Chinese/Indian inventors results in 1.605 additional patent grants. This estimate aligns well with, for instance, Amgen in 2015. That year Amgen filed for 420 LCAs, 80 of which led to H1B visas. In 2015, 80% of H1Bs were granted to Chinese/Indians,[20] suggesting that 64 Chinese/Indian nationals working for Amgen received H1B visas. We observe that Amgen filed 71 herbal patents that year, nine of which were granted. This pattern suggests that hiring 10 additional Chinese/Indian inventors results in 1.406 additional herbal patent grants, a number similar to the 1.605 obtained above. Details appear in the Appendix (Section 2).

Placebo test. Serial correlation in the treatment variable across years may bias standard errors in difference-in-differences estimates, causing us to underestimate the standard errors and to over-reject the null hypothesis. We follow the block bootstrapping suggestions in Bertrand, Duflo, and

[20] https://www.recode.net/2017/4/13/15281170/china-india-tech-h1b-visas

Mullainathan (2004) and Chetty, Looney, and Kroft (2009) to run a nonparametric permutation test to study whether our estimates suffer from such biases. Intuitively, the permutation test calculates the probability that we will see an equally large effect size when the treatment groups and treatment periods are randomly selected while preserving the assignee-level correlation structure. Figure A4 in the Appendix plots the cdf of the placebo estimates and a vertical line corresponding to the size of our DD coefficient. As this figure shows, we observe coefficients as large as the one in the fully specified model in Table 2 (i.e., 0.044) less than five percent of the time, boosting our confidence in the results.

Inventors' educational backgrounds. An empirical concern about using algorithms that code ethnicities using names is whether the ethnicities of multicultural individuals are identified correctly. Inventors' backgrounds can provide information about whether herbal-patent inventors are more likely to be first-generation migrants. We randomly sampled 552 inventors from the Chinese/Indian population and searched for their educational histories on LinkedIn. To do so, we searched LinkedIn for the inventors' and assignees' names. If we found a profile that (1) matched the inventor name and (2) matched the assignee of interest (3) close in time to the period when the patent application was submitted, we coded a search as successful. We matched 84 profiles (15% of our random sample) but dropped 20 individuals who did not list educational credentials. (See Tables A14–A16 in the Appendix). Approximately one-third of the sample was educated solely in India; a similar fraction was educated solely in the United States. About 20% were educated in China before moving to the United States to study and/or work. The remaining inventors were educated solely in China (9%) or educated in both India and the United States (3%). In summary, a disproportionate fraction of Chinese/Indian inventors who filed herbal patents were educated in China or India, indicating that they are indeed first-generation migrant inventors.

Clustering and alternate specifications. Our analysis pertinent to Hypotheses 1 and 2 is clustered at the assignee level because we believe that error terms for patents with the same assignee will be correlated. For instance, company-specific policies may affect the proportion of ethnic Chinese/Indian inventor names on a given company's herbal patents. An alternative and broader level at which to cluster would be the patents' IPC class, but doing so would reduce the effective number of clusters to <40, which might be too few for unbalanced panels (Cameron and Miller, 2015). Thus, we consider the assignee level appropriate to cluster standard errors. Our results (reported in appendix Table A8) are robust to nonlinear count models (Poisson regression, Poisson QML) and to nonlinear count models that explicitly allow for over-dispersion (negative binomial regressions).

Differences in investments in patent quality. Differences in incentives and resources between capped and exempt assignees raise possible endogeneity concerns. It is possible that capped firms are more likely to benefit from herbal patents, and that they also have greater resources (better knowledge workers and patent lawyers) with which to obtain granted patents. Our use of assignee fixed effects mitigates this concern. In addition, a *t*-test on the different means of grant probabilities between capped and exempt companies returns a difference of 0.039 with a standard deviation of 0.101, suggesting that the grant probabilities of the two groups do not significantly differ.

Value of herbal remedies to the western bio-pharma industry. As for whether herbal remedies are valuable to the western bio-pharma industry, a few stylized facts will shed light. The U.S. herbal-remedies market was worth $5.4 billion in 2016 and is forecasted to reach $6.6 billion by 2021 (Mintel, 2016). Well-known products in this segment include Metamucil (Procter & Gamble), Benefiber (GlaxoSmithKline), and Fibercon (Pfizer) (Euromonitor, 2016). Within western scientific research more generally, herbal and natural ingredients are cited as key sources for drug discovery (Doak *et al.*, 2014). Prior literature documents that, between 1981 and 2014, at least 33 percent of all

newly introduced chemical entities (NCEs) were natural-product-derived (Newman and Cragg, 2007). Figure A1 in the Appendix plots counts of articles on herbal remedies in all journals in PubMed and in selective journals like *Science*, *Nature,* and the *New England Journal of Medicine*.

Value created for firms. We next provide evidence that herbal patents filed by first-generation ethnic migrant inventors create value for firms (measured using patent citations). Our secondary analysis (reported in Table A6 in the Appendix) shows that citations of new-herb patents filed by capped firms increased by 91 percent during the visa shock. Capped firms experienced a supply shock of first-generation ethnic migrants during the visa shock. Given our prior finding that patents with new herbs are more likely to be filed by ethnic migrants, this pattern suggests that herbal patents filed by first-generation ethnic migrant inventors create value for firms.

Discussion and Conclusion

We studied the role of first-generation ethnic migrant inventors in cross-border transfer of knowledge previously locked within the cultural contexts of their home regions. We exploit an exogenous supply shock to U.S. immigration and a list of patenting entities excluded from the shock to present robust econometric results. We also find that ethnic migrant inventors are more likely to engage in reuse of their prior knowledge, whereas knowledge recombination is more likely to be pursued by teams comprising inventors from other ethnic backgrounds.

Contributions of our study

Our results contribute to several literatures, including those on skilled migration, ethnic migration, inventor mobility and knowledge flows, and the microfoundations of knowledge recombination. Like Jensen and Szulanski (2004), who argued that institutional distance increases the stickiness of knowledge and impedes its transfer, we argue that knowledge can be locked within the cultural and linguistic context where it was originally produced. We add to this literature by showing that hiring skilled ethnic migrants can help firms appropriate, and subsequently recombine, knowledge

previously locked in the home regions of ethnic migrants.

Our results contribute to the literature on skilled migration. Recent research and policy debate (Kerr and Lincoln, 2010; Kerr *et al.*, 2012; Doran *et al.*, 2016) have focused on the job-creation effects of the H1B program.[21] We sidestep that debate to highlight the role of migrant inventors in transferring across borders knowledge previously locked within their home regions. Contrary to the prevailing assumption that skilled migrants resemble local knowledge workers (and thus might displace them), our paper implies that skilled ethnic migrants can *differ* from locals with regard to the knowledge they bring to a host firm. This finding suggests that debate on skilled immigration should consider knowledge-transfer and knowledge-recombination effects with and without skilled migration. We also contribute to the skilled-migration literature by highlighting the role of first-generation ("new") migrants in knowledge transfer across borders.

There is also an emerging literature on the role of ethnic inventors and Diaspora in facilitating knowledge transfer (Saxenian, 2000; Kerr, 2008; Nanda and Khanna, 2010; Agrawal *et al.*, 2011; Foley and Kerr, 2013; Almeida, Phene, and Li, 2014; Choudhury 2015).[22] We contribute to it by studying the roles of ethnic and non-ethnic inventors in knowledge recombination, a topic that has not been fully explored in the previous literature. Like Freeman and Huang (2014), who find that diversity in author ethnicity is related to more citations and a higher impact factor, our results suggest that knowledge recombination is partly driven by mixed teams.

Our results contribute to the strategy literature on the micro-foundations of knowledge recombination (Allen, 1977; Fleming 2001) by heeding calls to elucidate the micro-foundations of

[21] Kerr and Lincoln (2010) find that changes in H1B admission levels influence the rate of Indian and Chinese patenting in cities and at firms dependent on the program. Kerr *et al.* (2012) find increases in firms' employment of skilled immigrants to be related to overall employment of skilled workers. But Doran *et al.* (2016) find that H1B visas crowd out employment of other workers.

[22] Kerr (2008) notes that ethnic scientific networks are central to short-term technology transfer from the United States. Agrawal *et al.* (2011) find that inventors who work for multinational firms in India cite the Indian Diaspora more frequently than do counterparts employed by the same firms in other countries. Almeida *et al.* (2014) find evidence of intra-ethnic citations in the U.S. semiconductor industry.

innovation within firms (e.g., Felin and Foss, 2005). The recent literature in this area includes the work of Gruber *et al.* (2013) on how inventors' individual characteristics (e.g., their educational backgrounds and whether they are scientists or engineers) affect the breadth of their technological recombinations. Other work (Fleming *et. al.*, 2007; Paruchuri and Awate, 2017) studies the effect of individual inventors' network positions on their ability to engage in recombination.[23] Our findings contribute to this literature by suggesting that, though ethnic migrant inventors may transfer knowledge from their home regions to western firms and reuse it, recombination is apt to be performed by non-ethnic inventors. This scenario indicates a complementary relationship between the ethnic migrant inventor and the non-ethnic inventor, an insight that recalls the literature on concurrent sourcing of complementary components for knowledge recombination (Parmigiani and Mitchell, 2009; Hess and Rothaermel, 2011).[24] Our findings also speak to a mechanism that Nerkar (2003) calls "temporal exploration" whereby firms create value by combining new knowledge with time-honored knowledge.

Generalizability of our results and boundary conditions

External validity is among our study's limitations. To explore the generalizability of our results, we did a comprehensive search of the literature in economic history and migration and profiled nine qualitative examples pertinent to the phenomenon of interest (Table A1 in the Appendix). They include the example of Italian migrants' transfer of knowledge specific to the food and fashion industries to the United States and Australia. As we document, Italians resisted assimilation in general, and specifically resisted "Americanizing" their cooking habits. Italian migrants' transfer of knowledge and ingredients to the United States gave way to subsequent recombination of

[23] Fleming, Mingo, and Chen (2007) study the brokerage positions of individual inventors; Paruchuri and Awate (2017) study the reach of inventors in intra-firm networks and their ability to span structural holes. Other papers in this literature include Nerkar and Paruchuri, 2005; Audia and Goncalo, 2007; and Tzabbar, 2009.

[24] Our results are closely related to those of Hess and Rothaermel (2011), who build on Arora and Gambardella (1990) by arguing that star scientists can link a firm to complementary, non-redundant knowledge at other organizations. Our insights contribute to the broader literature on intra-firm knowledge recombination (Carnabuci and Operti, 2013; Karim and Kaul, 2014).

knowledge: firms such as Campbell's and Heinz marketed shelf-stable versions of Italian cuisine, including such dishes as spaghetti in tomato sauce (Levenstein, 1985). In doing so, the American firms applied their own knowledge of processing and packaging food to classic Italian cooking. We also profile the example of British migrants to Italy who transferred back home their tacit know-how about operating silk machines (Cipolla, 1972). A further example is the transfer of accounting practices by Indian migrants from Gujarat to South Africa.

To reiterate, the key assumption underlying our phenomenon of interest is that, ex-ante, knowledge was initially *locked* in the cultural context of ethnic migrants' home region, due to causal ambiguity, and/or unprovenness of using the knowledge in a different context, and/or concentration in the home region of experts with the requisite tacit know-how. This boundary condition implies that, over time, if both the 'know-why' and the 'know-how' could be codified in the host region, such knowledge could be unlocked and could transfer to a new geography. As an example, Florida and Kenney (1991) study Japanese automotive assembly plants in the United States and conclude that Japanese production practices can be uncoupled from Japanese culture and transferred abroad. Bikard and Marx (2018) have also shown that the same knowledge might be *simultaneously* discovered across two geographical contexts, thus circumventing the need for transfer. Future research could further explore boundary conditions in other settings, such as restaurants (ethnic migrant chefs and sommeliers).

Other limitations and directions for future research

We capture the effects of immigration via the marginal H1B visa candidate, a highly skilled individual. More general increases in immigration may have different impacts on the transfer, codification, and recombination of knowledge previously locked in migrants' home regions.

Beyond the cultural dimension of geographic context, future research can explore how region-specific institutional factors influence why knowledge is locked in a particular geographic

context. It can also explore the role in cross-border knowledge transfer of ethnic scientists who are temporary migrants. An example is A.Q. Khan of Pakistan, a temporary migrant scientist who arguably transferred knowledge of centrifuge technology to North Korea.[25] Future work could also study whether ethnic migrants codify such knowledge in forms other than patenting, i.e., academic publications and business practices, and how the value of recombining knowledge from non-western settings with existing western knowledge should be shared between ethnic groups. An example of such a project is the Amazon Third Way initiative, an effort by the World Economic Forum to design and deploy the Amazonian Bank of Codes, an open digital platform that will map the biological assets of the Amazon and provide a global marketplace for such knowledge. In the broader strategy literature, scholars could study the role of skilled ethnic migrants in cross-border transfer of knowledge underlying cultural goods and services.[26]

We began this study by asking whether ethnic migrant inventors differ from locals in how they contribute to innovation at their host firms. Our research outlines at least one dimension along which ethnic migrants differ from non-ethnic knowledge workers, i.e., the knowledge they bring to the firm. Our research suggests that ethnic migrant inventors and non-ethnic inventors play different roles vis-à-vis knowledge reuse and knowledge recombination. Our results have managerial implications for firms engaged in R&D and cross-border sourcing of ideas. They also have implications for policy pertaining to high-skilled immigration and the effectiveness of temporary worker programs like the H1B.[27] In conclusion, our study suggests that a focus on whether migrants create or displace local jobs is too narrowly focused; social planners should consider the loss to cumulative knowledge production if western countries restrict their intake of skilled ethnic migrants.

---

[25] Source: https://qz.com/1080927/did-pakistan-help-north-korea-develop-nuclear-weapons-india-us-japan-want-to-know/

[26] There is a rich strategy literature on cultural goods (e.g., Lampel, Lant, and Shamsie, 2000) but little empirical work linking migration of ethnic knowledge workers to the spread of cultural goods across borders.

[27] Source: https://www.wsj.com/articles/indian-workers-in-u-s-fear-trump-h-1b-visa-crackdown-1488191404

References

Agrawal A, Kapur D, McHale J, Oettl A. 2011. Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *Journal of Urban Economics* **69**(1): 43–55.

Ahuja G, Morris Lampert C. 2001. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal* **22**(6–7): 521–543.

Allen TJ. 1977. *Managing the flow of technology: technology transfer and the dissemination of technological information within the R&D organization*, 1st pbk. print. MIT Press: Cambridge, Mass.

Almeida P, Kogut B. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management science* **45**(7): 905–917.

Almeida P, Phene A, Li S. 2014. The influence of ethnic community knowledge on Indian inventor innovativeness. *Organization Science* **26**(1): 198–217.

Amabile T, Kramer S. 2011. *The progress principle: Using small wins to ignite joy, engagement, and creativity at work*. Harvard Business Press.

Ambekar A, Ward C, Mohammed J, Male S, Skiena S. 2009. Name-ethnicity Classification from Open Sources. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09. ACM: New York, NY, USA: 49–58. Available at: http://doi.acm.org/10.1145/1557019.1557032.

Arora A, Gambardella A. 1990. Complementarity and external linkages: the strategies of the large firms in biotechnology. *The journal of industrial economics* : 361–379.

Audia PG, Goncalo JA. 2007. Past success and creativity over time: A study of inventors in the hard disk drive industry. *Management Science* **53**(1): 1–15.

Autor DH. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of labor economics* **21**(1): 1–42.

Belfanti C. 2004. Guilds, patents, and the circulation of technical knowledge: Northern Italy during the early modern age. *Technology and culture* **45**(3): 569–589.

Berry H, Guillén MF, Zhou N. 2010. An institutional approach to cross-national distance. *Journal of International Business Studies* **41**(9): 1460–1480.

Bertrand, M., Duflo, E., Mullainathan, S. 2004. How much should we trust differences-in-differences estimates?. *The Quarterly journal of economics*, **119**(1): 249-275.

Bikard, M. and Marx, M., 2018. Hubs As Lampposts: Academic Location and Firms' Attention to Science.

Borjas GJ, Doran KB. 2012. The collapse of the Soviet Union and the productivity of American mathematicians. *The Quarterly Journal of Economics* **127**(3): 1143–1203.

Breschi S, Lissoni F. 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of economic geography* **9**(4): 439–468.

Burt RS. 1992. *Structural holes: the social structure of competition*. Harvard University Press: Cambridge, Mass.

Cameron AC, Miller DL. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* **50**(2): 317–372.

Cameron AC, Trivedi PK. 2010. *Microeconometrics using stata*. Stata press College Station, TX, 2.

Carnabuci G, Operti E. 2013. Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strategic Management Journal* **34**(13): 1591–1613.

Chetty R, Looney A, Kroft K. 2009. Salience and taxation: Theory and evidence. *American economic review* **99**(4): 1145–77.

Choudhury P. 2015. Return migration and geography of innovation in MNEs: a natural experiment of knowledge production by local workers reporting to return migrants. *Journal of Economic Geography* **16**(3): 585–610.

Cipolla CM. 1972. The diffusion of innovations in early modern Europe. *Comparative Studies in Society and History* **14**(1): 46–52.

Cohen WM, Levinthal DA. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly* : 128–152.

Connelly BL, Certo ST, Ireland RD, Reutzel CR. 2011. Signaling theory: A review and assessment. *Journal of management* **37**(1): 39–67.

Cowan R, Foray D. 1997. The economics of codification and the diffusion of knowledge. *Industrial and corporate change* **6**(3): 595–622.

Dasgupta P, David PA. 1994. Toward a new economics of science. *Research policy* **23**(5): 487–521.

Doak BC, Over B, Giordanetto F, Kihlberg J. 2014. Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chemistry & biology* **21**(9): 1115–1142.

Doran K, Gelber A, Isen A. 2016. *The effects of high-skilled immigration policy on firms: Evidence from H-1B visa lotteries*. National Bureau of Economic Research.

Euromonitor. 2016. Herbal/traditional products in the US. Available at: http://www.portal.euromonitor.com.ezp-prod1.hul.harvard.edu/portal/analysis/tab [15 March 2017].

Felin T, Foss NJ. 2005. *Strategic organization: A field in search of micro-foundations*. Sage Publications London, Thousand Oaks, CA and New Delhi.

Fleming L. 2001. Recombinant uncertainty in technological search. *Management science* **47**(1): 117–132.

Fleming L, Mingo S, Chen D. 2007. Collaborative brokerage, generative creativity, and creative success. *Administrative science quarterly* **52**(3): 443–475.

Florida R, Kenney M. 1991. Organisation vs. culture: Japanese automotive transplants in the US. *Industrial Relations Journal* **22**(3): 181–196.

Foley CF, Kerr WR. 2013. Ethnic innovation and US multinational firm activity. *Management Science* **59**(7): 1529–1544.

Franzoni C, Scellato G, Stephan P. 2014. The mover's advantage: The superior performance of migrant scientists. *Economics Letters* **122**(1): 89–93.

Freeman RB, Huang W. 2014. Collaboration: Strength in diversity. *Nature News* **513**(7518): 305.

Ganguli I. 2015. Immigration and Ideas: What Did Russian Scientists "Bring" to the United States? *Journal of Labor Economics* **33**(S1): S257–S288.

Ghemawat P. 2001. Distance still matters. *Harvard business review* **79**(8): 137–147.

Gruber M, Harhoff D, Hoisl K. 2013. Knowledge recombination across technological boundaries: Scientists vs. engineers. *Management Science* **59**(4): 837–851.

Hambrick DC, Macmillan IC. 1985. Efficiency of product R&D in business units: The role of strategic context. *Academy of Management Journal* **28**(3): 527–547.

Henderson RM, Clark KB. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly* : 9–30.

Hess AM, Rothaermel FT. 2011. When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. *Strategic Management Journal* **32**(8): 895–909.

Holmström B. 1999. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies* **66**(1): 169–182.

Hornung E. 2014. Immigration and the diffusion of technology: The Huguenot diaspora in Prussia. *American Economic Review* **104**(1): 84–122.

Jensen R, Szulanski G. 2004. Stickiness and the adaptation of organizational practices in cross-border knowledge transfers. *Journal of international business studies* **35**(6): 508–523.

Johnson CY. 2017, February 1. Big pharma depends on immigrants. It kept quiet about Trump's travel ban. *Washington Post*. Available at: https://www.washingtonpost.com/news/wonk/wp/2017/02/01/big-pharma-depends-on-immigrants-it-kept-quiet-about-the-travel-ban/.

Karim S, Kaul A. 2014. Structural recombination and innovation: Unlocking intraorganizational knowledge synergy through structural change. *Organization Science* **26**(2): 439–455.

Kerr SP, Kerr W, Özden Ç, Parsons C. 2016. Global talent flows. *Journal of Economic Perspectives* **30**(4): 83–106.

Kerr SP, Kerr WR, Lincoln WF. 2015. Skilled immigration and the employment structures of US firms. *Journal of Labor Economics* **33**(S1): S147–S186.

Kerr WR. 2008. Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics* **90**(3): 518–537.

Kerr WR, Lincoln WF. 2010. *The supply side of innovation: H-1B visa reforms and US ethnic invention*. National Bureau of Economic Research. Available at: http://www.nber.org.ezp-prod1.hul.harvard.edu/papers/w15768.

Lampel J, Lant T, Shamsie J. 2000. Balancing act: Learning from organizing practices in cultural industries. *Organization science* **11**(3): 263–269.

Levenstein H. 1985. The American response to Italian food, 1880–1930. *Food and Foodways* **1**(1–2): 1–23.

Lewin AY, Massini S, Peeters C. 2009. Why are companies offshoring innovation? The emerging global race for talent. *Journal of International Business Studies* **40**(6): 901–925.

Michel S, Witte J. 2014. *Immigrants Working for U.S. Pharmaceuticals*. George Mason University, Fairfax, VA. Available at: http://s3.amazonaws.com/chssweb/documents/16298/original/Immigrants_in_the_Pharma_Industry_Institute_for_Immigration_Research_GMU.pdf?1407181243.

Mintel. 2016. Homeopathic and Herbal Remedies - US - November 2016 - Market Research Report. Available at: http://academic.mintel.com.ezp-prod1.hul.harvard.edu/display/765799/ [15 March 2017].

Nanda R, Khanna T. 2010. Diasporas and domestic entrepreneurs: Evidence from the Indian software industry. *Journal of Economics & Management Strategy* **19**(4): 991–1012.

Nelson RR, Winter SG. 1982. *An Evolutionary Theory of Economic Change*. SSRN Scholarly Paper, Social Science Research Network, Rochester, NY. Available at: https://papers.ssrn.com/abstract=1496211.

Nerkar A. 2003. Old is gold? The value of temporal exploration in the creation of new knowledge. *Management Science* **49**(2): 211–229.

Nerkar A, Paruchuri S. 2005. Evolution of R&D capabilities: The role of knowledge networks within a firm. *Management science* **51**(5): 771–785.

Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last 25 years. *Journal of natural products* **70**(3): 461–477.

Oettl A, Agrawal A. 2008. International labor mobility and knowledge flow externalities. *Journal of international business studies* **39**(8): 1242–1260.

Parmigiani A, Mitchell W. 2009. Complementarity, capabilities, and the boundaries of the firm: the impact of within-firm and interfirm expertise on concurrent sourcing of complementary components. *Strategic Management Journal* **30**(10): 1065–1091.

Paruchuri S, Awate S. 2017. Organizational knowledge networks and local search: The role of intra-organizational inventor networks. *Strategic Management Journal* **38**(3): 657–675.

Polanyi M. 1966. *The tacit dimension*, 1st ed. Terry lectures. Doubleday: Garden City, N.Y.

Rosenkopf L, Almeida P. 2003. Overcoming local search through alliances and mobility. *Management science* **49**(6): 751–766.

Saxenian A. 2000. Silicon Valley's New Immigrant Entrepreneurs. Available at: https://escholarship.org/uc/item/88x6505q.

Schumpeter JA. 1939. *Business cycles*. McGraw-Hill New York, 1.

Simonsohn U. 2017. Two-lines: The first valid test of U-shaped relationships.

Singh J. 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Management science* **51**(5): 756–770.

Song J, Almeida P, Wu G. 2003. Learning–by–Hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Science* **49**(4): 351–365.

Szulanski G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic management journal* **17**(S2): 27–43.

Szulanski G. 2002. *Sticky knowledge: Barriers to knowing in the firm*. Sage.

Tzabbar D. 2009. When does scientist recruitment affect technological repositioning? *Academy of Management Journal* **52**(5): 873–896.

Von Hippel E. 1994. "Sticky information" and the locus of problem solving: implications for innovation. *Management science* **40**(4): 429–439.

Vroom VH. 1964. *Work and motivation*. Wiley: Oxford, England.

Table 1. Summary statistics

Capped	(1) Normal visa cap		(2) Increased visa cap		(3) Diff (2) – (1)	
	mean	sd	mean	sd	b	Se
Patent counts	0.06	0.29	0.21	0.55	0.15	(0.02)
Recombined	0.02	0.14	0.07	0.34	0.05	(0.01)
New inventors	0.11	0.66	0.39	1.42	0.28	(0.04)
New ethnic inventors	0.01	0.14	0.05	0.34	0.04	(0.01)
Ethnic inventors	0.35	0.85	0.38	0.83	0.04	(0.07)
Observations	5470		1384		6854	
Exempt	(4) Normal visa cap		(5) Increased visa cap		(6) Diff (5) – (4)	
	mean	sd	mean	sd	b	Se
Patent counts	0.04	0.21	0.12	0.41	0.08	(0.02)
Recombined	0.01	0.10	0.05	0.28	0.04	(0.02)
New inventors	0.10	0.58	0.24	1.00	0.14	(0.05)
New ethnic inventors	0.02	0.22	0.04	0.29	0.02	(0.02)
Ethnic inventors	0.57	1.07	0.56	1.08	-0.01	(0.22)
Observations	1788		356		2144	
Difference-in-Difference					Diff=in-Diff (6) – (3)	
					b	Se
Patent counts					0.07	(0.02)
Recombined					0.01	(0.01)
New inventors					0.14	(0.05)
New ethnic inventors					0.02	(0.01)
Ethnic inventors					0.05	(0.20)
Observations					8998	

Note: Standard errors appear in parentheses. Observations are at the assignee-year level, for all years that an assignee (firm or university) was in operation. We use the *tsfill* command in Stata to fill in missing assignee-year pairs. The variable "Patent counts" refers to the number of herbal patents filed by an assignee in a given year. The variable "Recombined" measures the number of recombined patents at the assignee-year level. In regressions, we use "Recombined" as an indicator for whether the patent contains synthetic compounds as well as herbal ingredients. The variables "New inventors" and "New ethnic inventors" represent the number of inventors in a firm year who file their first patent, and the same number for inventors with Chinese/Indian names. The variable "Herb count" counts the average number of herbs on a patent.

Table 2. The effect of a visa shock on herbal patents

	(1)	(2)	(3)	(4)
	Dependent variable: log(1+patent count)			
Capped x Shock	0.044	0.038	0.042	0.044
	(0.017)	(0.017)	(0.017)	(0.018)
Capped	0.010	0.016	0.009	
	(0.005)	(0.005)	(0.005)	
Shock	0.048	0.052		
	(0.014)	(0.014)		
Constant	0.030	0.042	0.000	0.027
	(0.004)	(0.006)	(0.010)	(0.010)
Controls	N	Y	Y	Y
Year FE	N	N	Y	Y
Assignee FE	N	N	N	Y
Observations	8998	8998	8998	8998
Adjusted R^2	0.030	0.059	0.068	0.072

Note: Standard errors appear in parentheses, clustered at the assignee level. Observations are at the assignee-year level, for all years that an assignee (a firm or university) was in operation. We use the *tsfill* command in Stata to fill in missing assignee-year pairs. The assignee-year-level dataset is thus an unbalanced panel consisting of 8,998 observations (an average of 22.4 years of observations for 401 assignees). The dependent variable is the log of the number of herbal patents filed by an assignee in a given year. *Capped* is an indicator for whether the assignee is subject to the visa cap; *Shock* is an indicator for years 1999 through 2003. Controls include the fraction of inventors who are Chinese/Indian, assignee age, number of inventors, and the number of Chinese/Indian inventors. Percentage increases are calculated as $100 \cdot (e^{\beta} - 1)$. In the base case, we use an OLS specification. For models with assignee fixed effects, we use an xtreg specification. The dependent variable, number of herbal patents granted to an assignee in a given year, has many zeros; and the variance of this variable is larger than its mean. Consequently, we investigate whether a Poisson regression is appropriate for this data. Following Cameron and Trivedi (2010), we check whether the outcome variable looks like a Poisson-distributed random variable. We see a slightly higher probability of zeros in the observed outcome than the Poisson distribution would predict (0.0089), but no difference in predicted counts of zeros using Stata's countfit command. The contribution of zeros to the Pearson Chi-Square statistic is 0.001, further showing that our data does not suffer from over inflation of zeros. We also ran robustness checks using a Negative Binomial, Poisson and Poisson QML specification. Alternatively, in the Appendix (Table A18) we present a DD specification using Post2004 (tightening immigration policies) as the treatment period, which yields a negative and significant coefficient (β=-0.036, p<0.001).

Table 3. Probability of recombination across ethnicities

	(1)	(2)	(3)	(4)
	Dependent variable: Recombined			
Fully Ethnic	-0.346		-0.609	
	(0.154)		(0.318)	
Capped x Shock		-0.313	-0.243	
		(0.136)	(0.132)	
Capped x Shock x Fully Ethnic			-0.897	
			(0.218)	
Fully Ethnic x Shock			0.872	
			(0.198)	
Fully Ethnic x Capped			0.293	
			(0.288)	
Non-ethnic				0.360
				(0.155)
Mixed Teams				0.315
				(0.162)
Constant	-0.995	0.481	0.425	-1.272
	(4.688)	(0.186)	(0.185)	(4.737)
Controls	Yes	Yes	Yes	Yes
Assignee FE	Yes	Yes	Yes	Yes
Application Year FE	Yes	Yes	Yes	Yes
Observations	758	758	758	758
Adjusted R^2	0.062	0.073	0.078	0.062

Note: Standard errors appear in parentheses, clustered at the assignee level. Observations are at the patent level. We retain only herbal patents filed by U.S. assignees in operation during the visa-shock period. The dependent variable is an indicator for whether the patent contains synthetic compounds. *Fully Ethnic* is an indicator for patents with only ethnic inventors. *Capped x Shock* is an indicator denoting whether the patent was filed by a cap-subject firm between 1999 and 2003 (the visa-shock period), and *Capped x Shock x Fully Ethnic* interacts this with an indicator for Full Ethnic teams. *Non-Ethnic* and *Mixed teams* are indicators for patents listing no inventors with Chinese/Indian names and for patents with some but not all Chinese/Indian inventors. The omitted group for column (1) is patents with any non-Chinese/Indian inventors. In column 4, the omitted group is patents by teams composed entirely of Chinese/Indian inventors. Controls include assignee age at the time of application, number of claims listed on the patent, whether the patent had any new inventors, average prior co-inventor centrality, and average prior exposure to ethnic co-inventors. In the base case, we use an OLS specification (given that all models include assignee fixed effects, we use an xtreg specification).

Table 4. Recombination probabilities over time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Recombined		Recombined (non-ethnic inventors)		Recombined (mixed team)		Recombined (ethnic inventors)	
Time since herb introduced	0.0053	0.0188	0.0033	0.0186	0.0029	-0.0027	-0.0009	0.0029
	(0.0012)	(0.0031)	(0.0011)	(0.0029)	(0.0007)	(0.0020)	(0.0002)	(0.0008)
Time since herb introduced squared		-0.0004		-0.0005		0.0002		-0.0001
		(0.0001)		(0.0001)		(0.0001)		(0.0000)
Constant	0.1655	0.0893	0.2605	0.1737	-0.1348	-0.1033	0.0398	0.0188
	(0.0210)	(0.0229)	(0.0167)	(0.0208)	(0.0125)	(0.0137)	(0.0043)	(0.0057)
Herb FE	Y	Y	Y	Y	Y	Y	Y	Y
Patent Level Controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	4849	4849	4849	4849	4849	4849	4849	4849
Adjusted R^2	0.020	0.023	0.083	0.088	0.176	0.178	0.062	0.065

Notes: Standard errors appear in parentheses, clustered at the herb level. Each observation is an herb-patent pair, for all herbs and for all patents. For example, a patent with three herbs would count as three data points. We have 758 patents (note that sample size for Table 2 is 758) and end up with 4849 herb-patent pairs. The variable "Time since herb introduced" for an herb-patent pair is calculated by subtracting the patent-application year from the first year we observe the focal herb in any patent, across all patents. Our dependent variable measures, for each herb-patent pair, whether the patent uses synthetic compounds. Because we are tracking, for a given herb, the likelihood of recombination over time, we include herb fixed effects. We include patent-level controls to control for patent-level factors that may confound the relationship. The median herb is 17 years old, suggesting that a median herb is 20.4 percentage points more likely to be recombined than a newly introduced herb (from Column (2), 17*0.0188 - 17^2*0.0004=0.3196-0.1156=0.204).

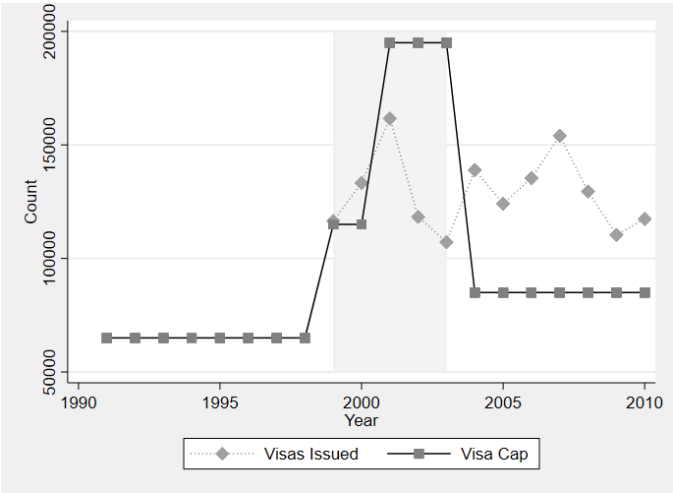
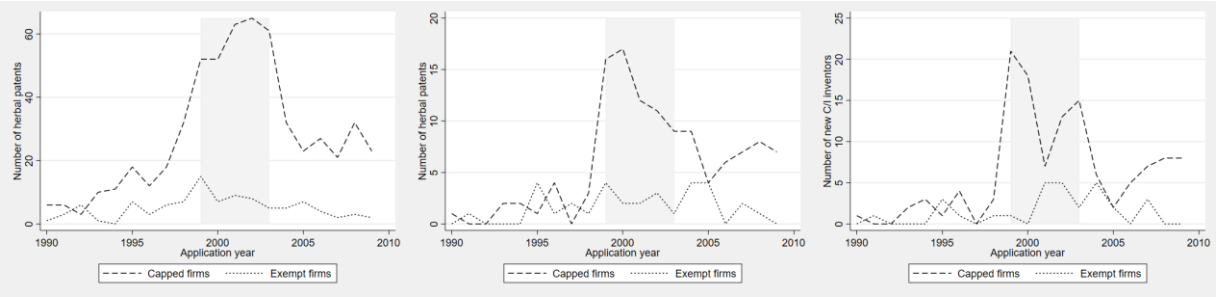


Figure 1. H1B visa cap over time

Notes: Figure 1 plots the H1B visa cap and visa issuances over time. The shaded area represents the period during which the American Competitiveness in the 21st Century Act (AC21) was in effect. In accordance with the AC21 Act, H1B visa quotas were raised between 1999 and 2003 and lowered starting in 2004. The American Competitiveness and Workforce Improvement Act (ACWIA) passed in 1998 increased the H1B visa cap from 65,000 to 115,000. The American Competitiveness in the 21st Century Act (AC21) passed in 2000 further increased the visa cap to 195,000. AC21 also retroactively increased the 1999 and 2000 quotas above the 115,000 cap set by the ACWIA. The actual H1B visa issuances in 1999-2003 were 116,513 (1999), 133,290 (2000), 161,643 (2001), 118,352 (2002), and 107,196 (2003), all greater than the pre-1999 cap of 65,000. Visa issuances post-2003 are above the visa cap, as AC21 created the cap-exempt group.



(a) All herbal patents *(b) Herbal patents by Chinese/Indians* *(c) New Chinese/Indian inventors*

Figure 2. Number of herbal patent grants and new inventors over time

Notes: Shaded areas represent the period during which the visa cap was increased due to the AC21 Act. We see a sharper increase in the number of herbal patents during the visa-shock period, and especially in herbal patents with ethnic inventors for capped firms, compared to exempt firms. The number of new Chinese/Indian inventors also rises significantly during the same visa-shock period, especially for capped firms.

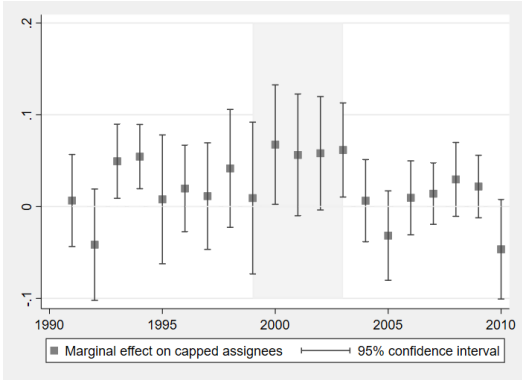


Figure 3. Estimated impact of visa shock on herbal patenting for capped versus exempt firms in years before, during, and after 1999-2003 shock

Notes: Standard errors are clustered at the assignee level. The shaded area represents the period during which the visa cap was raised by the AC21 Act. We estimate $\ln(1 + Patent\ Count_{jt}) = \lambda_t + Capped_j + \sum_\tau \delta_\tau(\lambda_t \times Capped_j)$ where we include dummies for capped firms ϕ_j, year fixed effects λ_t, and all interactions between year dummies and capped firms. Figure 3 plots the coefficients and confidence intervals for all interaction terms δ_τ.

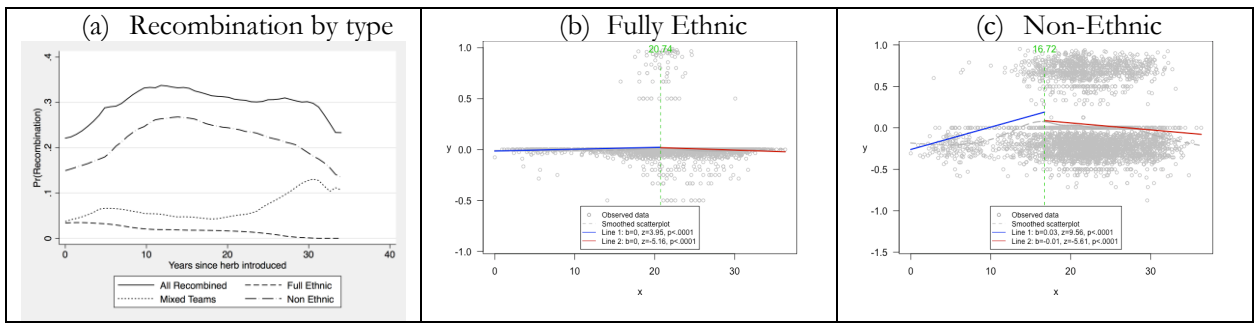


Figure 4. Recombination probabilities for herbs over time since their introduction to the U.S.

Notes: This figure visualizes results reported in Table 4 (i.e. recombination probabilities over time). Plots are generated via Kernel-weighted local polynomial regressions using Stata's *lpoly* command. The unit of analysis is at the herb-year level. The x-axis plots the time since an herb was introduced. The dependent variable is whether an herb was used in a patent for recombination at time t. Recombination is measured at the patent level, and can be a recombination using a synthetic compound, along with multiple herbs. Each point on the line thus corresponds to the probability of being recombined (by an ethnic group), for a given herb, t years after its introduction (i.e., $p(t) = E\left[\frac{\gamma_h(t)}{N_h(t)}\right] = \frac{1}{\sum_t N_h(t)} \sum_h \frac{\gamma_h(t)}{N_h(t)}$, where $N_h(t)$ is the number of patents using herb h at time t, and $\gamma_h(t)$ is the number of recombined patents (by an ethnic group) using herb h, at time t. The solid line in panel (a) represents recombination by any ethnic group, and therefore is the sum of all other lines. We see that patents overall and patents by non-Chinese/Indian teams exhibit an inverted-U-shaped pattern with respect to recombination probabilities. In panels (b) and (c), we estimate an interrupted regression with the break point determined by a "Robin Hood" algorithm as in Simonsohn (2017). Panel (b) plots the results for any recombination, and panel (c) plots the results for recombination by fully non-ethnic teams. An inverted U-shape relationship exists if the linear coefficient to the left of the breakpoint is positive and significant, and the linear coefficient to the right of the breakpoint is negative and significant. While not shown here, recombination by mixed teams is gradually increasing over time.

Table of Contents

List of Tables

List of Figures

Section 1. Generalizability

1.1 Qualitative Case Studies Related to Phenomenon of Interest

Table A1. *Qualitative Case studies*

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
1	Migration of Huguenots to Brandenburg-Prussia	Huguenots (French Protestants) migrated to Brandenburg-Prussia (Northern German states) from 1685 till 1789. The migration was motivated by religious persecution in France and religious protection policies in Brandenburg-Prussia.	Many Huguenots distinguished themselves as skilled artisans and craftsmen, especially cloth-workers, and bore specialized and "secret" knowledge related to these crafts. One account "provides a list of 46 professions introduced by Huguenots to Brandenburg, all of which were previously unknown to the country, most of them in the textile industries. One Huguenot carried with him the secret of dyeing fabrics in a special way, another brought the art of printing on cotton (Hornung 2014, 91, 93).	In one example, Huguenots transmitted knowledge about silk production into Brandenburg-Prussia (Hornung 2014, 94). This stimulated a surge in the planting and cultivation of one preexisting crop in Brandenburg-Prussia: mulberry plants. There was a general impression that silk worms cannot thrive in the cold temperatures of the northern German states, such as Prussia. To help overcome this, concurrent with the influx of Huguenot immigrants, rulers in the northern German states began to promote the planting of mulberry tree plantations in their territories, the crop of which was used to feed silkworms. The historical record notes that "it is schoolmasters who chiefly occupy themselves" with the planting of silkworms, as a means of "adding to [their] income." The rule of Brandenburg-Prussia, who had encouraged Huguenot immigration, is noted to have "ordered the cultivation of mulberry trees in schoolyards to feed the silkworms." The resultant quality of silk in the northern German countries has been noted as "remarkably white, and finer than that in the southern countries (Scientific American 1853).

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
2	Migration of Soviet mathematicians to the United States	Mathematicians from the Soviet Union migrated to the United States between 1990 and 2000, after the fall of the Soviet Union. In 1992, the Soviet Scientists Immigration Act was passed into law in the United States, which allocated visas to be given to Soviet Scientists to immigrate to the U.S. As Ganguli (2015) states, estimates from the 2000 Census suggest that close to 10,000 Russian scientists and engineers across many science and technology fields immigrated to the United States in the 1990s.	During the Soviet era, Soviet mathematicians worked in mathematical knowledge sub-fields that differed from those Americans worked in (Borjas and Doran 2012, 6-9). Borjas and Doran (2012) show that Russian mathematicians were ahead of the west in fields like partial differential equations and symplectic topology. Such knowledge was also arguably ex ante "locked" within the soviet context, prior to the migration of Soviet scientists to the United States. Borjas and Doran (2012) outline two reasons for why such knowledge was locked within the Soviet context. First of all, as Abramitzky and Sin (2011) report, the translation rate of hard-science Eastern Bloc books into English was extremely low. Secondly, Borjas and Doran (2012) cite examples from Tybulewicz (1970) to state that even if translations of Soviet academic work were available, the knowledge was often not transferred to American academics, because they did not read the translations of Soviet scientific work.	An influx of Soviet mathematicians (roughly 300 during the 1990s) immigrated to the U.S. after the Soviet regime waned and collapsed in the early 1990s (Borjas 2014, 183). While they competed with American mathematicians for jobs and publication space in journals, they also collaborated and helped American mathematicians solve formerly intractable problems (Borjas and Doran 2012, 26, 11). One study cited news coverage from the time (from the New York Times) which described how "[American mathematician] Dr. Diaconis said he recently asked [Soviet mathematician] Dr. Reshetikhin for help with a problem that had stumped him for 20 years. 'I had asked everyone in America who had any chance of knowing' how to solve a problem . . . No one could help. But . . . Soviet scientists had done a lot of work on such problems. 'It was a whole new world I had access to,' Dr. Diaconis said. 'Together, we'll be able to solve the problem.

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
3	Migration of skilled artisans from Florence to Venice (driven by patent laws) and the influence of such migration on Silk Dyeing in Venice	Belfanti (2004) documents an interesting case study where migration of skilled artisans in Italy between the sixteenth and eighteenth century was driven by patent laws. As Belfanti (2004) documents, the city council and princes of Italian Republics such as the Venetian Republic used the attractiveness of assigning monopoly rights to intellectual property through patents, to attract skilled migrant workers from other regions, such as Florence. An interesting example related to this phenomenon dates back to the eighteenth century, when skilled silk dyers migrated from Florence to Venice.	While artisans from Venice possessed deep knowledge in making Brocade, they lacked knowledge in the area of silk dyeing. To transfer knowledge related to silk dyeing from Florence to Venice, as Belfanti (2004) documents, at the beginning of the eighteenth century, emissaries of Venice traveled to Florence to recruit silk dyers—an effort that succeeded in convincing dyers such as Cosmo Scatini, to migrate to Venice. Scatini was a Florentine dyer who knew the secret of dyeing silk black. Between 1727 and 1732 the Venetians also twice tried to bring experts in the making of silk veils from Bologna. To attract migrant artisans who possessed specialized knowledge to migrate to Venice, the Venetian Republic awarded them with patents. Cosmo Scatini, the migrant from Florence, was awarded with a patent for skill dyeing. Another migrant artisan craftsman who introduced the weaving of silk stockings on a frame was rewarded with a patent and a ten-year monopoly.	While the Venetian Republic used patents to attract skilled migrants from Florence and elsewhere, there was also a strong guild of local artisans in Venice, and the incentives of local artisans were geared towards working on the patented technology, at the end of the patent term. While the patents awarded to skilled migrant artisans granted them monopoly rights for a period of time, at the end of the fixed monopoly term, the migrant artisans were required to share their knowledge with the local guild of artisans. As Belfanti (2004) says, the Florentine artisan Cosmo Scatini, who obtained a patent for silk dyeing, applied to be enrolled in the dyers' guild once his privilege ran out, promising to teach the Venetian craftsmen the process. This transfer of knowledge often led to knowledge recombination. An example of such knowledge recombination relates to dyeing Black silks in Venice. Black was a symbol of superiority in Venice in the late 17th century (Bervegliari 1983, 176). When Cosmo Scatini transferred his secret knowledge of black silk dyeing to Venetians at the end of his patent term, the dyeing techniques changed in Venice and they were able to produce black silks (Bervegliari 1983, 176).

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
4	Migration of Italians to United States	As Choate (2008) documents, between 1880 and 1924, more than four million Italians immigrated to the United States, half of them between 1900 and 1910 alone—the majority fleeing grinding rural poverty in Southern Italy and Sicily. In the period 1880 to 1915, total Italian emigration has been estimated at 13 million, making it the largest emigration from any country in recorded history.	Natives of Campania and Sicily nurtured a cuisine of the tomato, onion, oil, cheese, and garlic. As Levenstein (1985) documents, the preparation and consumption of food was central to Italian family life. Italian women in particular "retained" knowledge about "community-borne recipes and instructions in cooking." (Levenstein 1985, 76, 80). Knowledge about *how* to cook with key ingredients such as these was tacit knowledge possessed by families, and was transferred to the United States by the migrants.	Italians in America resisted assimilation in general, and in particular resisted "Americanizing" their own cooking habits. A type of distinctive Italian cuisine in America thus took hold and key dishes became popular in American households after World War I. To bring basic Italian dishes and cooking methods to American households, such as "spaghetti in tomato sauce," brands such as Campbell's and Heinz marketed shelf-stable versions (Levenstein 1985, 79-81, 86). In doing so, they applied their own knowledge of processing and packaging food to traditional Italian recipes. As Levenstein (1985) documents, this recombination is in stark contrast to the negative perception towards ingredients such as tomatoes, traditionally harbored by American society. In the late 1880s, prior to the migration from Campania, the scientific community in the United States held a belief that tomatoes were carcinogenic, and were generally harmful due to the presence of "oxalic acid". The Italian migration, the transfer of knowledge related to Italian ingredients to the United States and the subsequent recombination of such knowledge went a long way to allay prior scientific "beliefs".

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
5	Ethnic migrants and their influence on the garment industry in New York	The New York garment industry has been shaped by waves of ethnic migrants from Germany, Ireland, Russia (Russian Jews), Italy, China, and Puerto Rico.	Till the late 1800s, different ethnicities migrating to New York transferred knowledge of their distinctive sartorial designs to their host region. As Bagger (1871) states, in the mid-1800s in New York, an individual's nationality could be determined by how he or she dressed. To quote the author, "It is curious to see such a heterogeneous crowd land. The Swedes are usually distinguished by their tanned-leather breeches and waistcoats, and their peculiar before mentioned exhalations; you can not miss the Irishman with his napless hat, worn coat, and corduroy trousers; the Englishman you know by his Scotch cap, clay pipe, and paper collar".	Among other scholars, Rantisi (2002) documents subsequent recombination of knowledge transferred by ethnic migrants, in the context of the New York garment district. The garment district in New York was a giant melting pot for ethnicities to work together and for knowledge recombination. As Rantisi (2002) states, while the origins of the New York garment industry can be traced to German Jews, migration of Italians in the 1880s and Russian Jews later led to knowledge recombination. In an online article, Dzvinka Stefanyshyn provides examples of ethnic migrant influence on knowledge transfer and knowledge recombination. To quote the author, "when the Italian and Jewish immigrants dominated clothing manufacturing businesses, certain aspects of style were altered. Up until the 1900s, women wore tight-fitting clothing, which included uncomfortable corsets and long, extravagant dresses. Turning away from that fashion, the workers focused on creating clothing that was much simpler and comfortable. This resulted in the introduction of much looser, natural – either woolen or linen – clothing".

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
6	Migrants from England to Italy (the silk machines example)	Temporary migration from England to Italy by migrants such as John Lombe in 1607, presumably for transferring knowledge related to "silk throwing".	In 1607 Vittorio Zonca published in Padua *Nuovo Teatro di Machine et Edificii* describing a machine to throw silk by water power. While this book was available in the UK, providing the know-why, it was only after a temporary migrant, John Lombe traveled to Italy and spent 2 years learning the know-how, the knowledge of silk throwing transferred to the UK (Cipolla 1972). To quote the author, "In 1607 Vittorio Zonca published in Padua his *Nuovo Teatro di Machine et Edificii* which included, among numerous engravings of various machines, the description of an intricate machine for throwing silk by water power in a large factory….notwithstanding the description by Zonca, the details of the mill were still considered state secret and Piedmontese laws regarded 'the disclosing or attempting to discover' anything relating to the making of the engines a crime punishable by death. The Piedmontese were no fools. G. N. Clark has pointed out that a copy of the first edition of Zonca's book had been on the open-access helves of the Bodleian Library from at least as early as 1620. Yet the English succeeded in building a mill for the throwing of silk only after John Lombe, during two years of industrial espionage in Italy, 'found means to see this engine so often that he made himself master of the whole invention and of all the different parts and motions" (Cipolla, 1972, page 47)	When Lombe returned from Italy to Britain, he brought with him first-hand know how of using of silk machines, and workers who could help set up the factories (Calladine 1993) to create British silk machines.

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
7	Italian migrants in Australia	As Castles (1992) documents, mass migration from Italy to Australia happened between 1950-1970; between 1951-1961, 18,000 new migrants arrived every year.	Genovesi (2000) documents the transfer of knowledge related to food and fashion by Italian migrants in Australia. To quote the author, "One needs to understand that Italian migrants needed particular foods for the practicing of social events, cultural and catholic religious rituals. For example, sugar almonds were needed for the rituals of the sacraments, dried fish for Good Friday, "la castagnata" (chestnuts) Christmas sweets, essence for cakes, alcohol such as wine and grappa. Amongst the Italians, there was little tolerance for mutton, fish and chips, or the meat pie". The author also documents transfer of knowledge across borders. To quote, "Many Italian immigrants felt the need to expand on the availability of vegetables and fruits and took matters into their own hands…..They had brought with them skills to make pasta, breads, sauces, preserves, cheeses, wines and pork sausages. In addition, as the immigrants arrived, so did particular foods, seeds, kernels and cuttings, mostly as contraband. In my conversations within the Italian community, it has been stated that these items were hidden in suitcases, coat linings, pockets and underwear." (Genovesi, 2000; page 8).	Over time, Italian food (and fashion) has experienced recombination in Australia. An example relates to the dish, Chicken Parmigiana. One account describes the dish as "An Italian name, but a bona fide Australian pub classic, the parmigiana started as an eggplant dish in Italy and has since evolved into a chicken schnitzel topped with an Italian-inspired tomato sauce and melted cheese." (source: http://www.cnn.com/travel/article/australian-food/index.html). Another source describes how the original recombination was within the Italian migrant communities in the United States and then the recombined dish became popular within Italian migrants in Australia. To quote, "Chicken Parmigiana has its origins in the United States, where it was popularized among Italian-American communities. Italian immigrants created the meal, which quickly became conceived as an authentically Italian dish. Of course, it does take its inspiration from Italy. Eggplant Parmigiana, or Mellenzana alla Parmigiana, is the original Italian recipe. Eggplants are lightly breaded, fried, topped with fresh tomato sauce and Parmesan cheese, and then baked. The switch to chicken in the United States might have been due to several reasons – Italian restaurant owners saw the American preference for meat over eggplant, Italian immigrant workers were able to afford meat now that they had higher paying jobs, or eggplants just weren't as common a produce in the United States." (source: https://www.montebene.com/blogs/blog-posts/58998467-the-story-behind-the-staple-chicken-parmigiana)

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
8	Migration of Indians to South Africa and East Africa and the impact of such migration on the transfer or accounting practices across borders	Indian merchants from Gujarat migrated to Natal, South Africa and to Kenya between 1875 and 1910 in search of better economic opportunities. This example documents the transfer of knowledge related to accounting practices from their home region to their host region, and subsequent recombination of such knowledge at their host region.	The migrants transferred knowledge related to an accounting practice called "Hundi" from Gujarat, India to Africa. A Hundi is most often translated as a "bill of exchange" or "promissory note." It functions as both a credit system and a remittance system, and historically has been linked to the Indian merchant community conducting trade in the Indian Ocean region. The bookkeeping system associated with the hundi exchange is known as *hundi khata*: a system which emphasizes "double-entry" accounting, and thus differs from the traditional English "bills book" system (Nigam 1986, 148, 156). The knowledge of this accounting practice was locked within the Gujarat region of India, because it was documented in vernacular text. The British Museum has referenced an autobiographical poem published in the 17th century by the Gujarati (Jain) poet and businessman Banarasidas as containing one of the earliest known references to hundi and Martine (2009) documents that the accounting practice was used exclusively within Gujarat because the creditors used the local vernacular language to write the promissory note.	Post the migration of Gujarati traders to Africa, the hundi system flourished in East Africa. European banks in the region also began to encourage the use of the "overdraft" as a medium for credit lending to "reliable" customers. The combination of the hundi and overdraft systems contributed to the development of a loosely-defined "chit" system in East Africa, as distinct from lending and accounting conventions popular in England at the time (Gregory 1993, 103). Prior research has documented that this recombined system allowed individuals to write checks or withdraw currency well beyond the amounts of their deposits on the understanding that, when able, they would make up the difference and pay a small interest on the overdrafts. As Gregory (1993) states, East Africa as early as 1907 was becoming known as 'the land of the chit.'" Notably, the English word "chit"—defined as "a signed voucher of a small debt"—itself originated from Hindi and Urdu, languages in Northern India (See Merriam-Webster, "Chit," accessed September 11, 2017, https://www.merriam-webster.com/dictionary/chit.)

#	Qualitative Example	Details of Migration	Knowledge Being Transferred Across Borders	Subsequent Recombination of Knowledge
9	Migration of Chinese to the United States and the creation of "American Chinese" food	Chinese migrants moved to the US because of the California Gold Rush in the 1850s. The early Chinese migrants were Cantonese from Guangdong area who entered primarily through port of San Francisco (source: The Search for General Tso). The Chinese Exclusion Act was signed in 1882, blocking Chinese immigration and naturalization, which led to attacks and threats on those already living in the US (Coe 2009). Chinese settlers began to move out of California and spread across the US to escape persecution and search for employment opportunities.	The documentary film The Search of General Tso documents the transfer of knowledge related to food from China to the U.S. When Chinese migrants first arrived to the U.S., Americans were both fascinated and repulsed by their food. After the signing of the 1882 Chinese Exclusion Act, many Chinese migrants were forced out of labor and had to be self-employed, occupying two main industries: laundry and food (source: The Search for General Tso). Chinese migrants began opening restaurants across the US. General Tso's chicken, arguably the most popular Chinese food dish in America, was brought to the U.S. from Taiwan, where former Hunan chef Peng Chang-kuei had fled after the Communist Revolution of 1949 (Bateman 2016). It was there that he created General Tso's chicken, inspired by the spicy and sour Hunan palate. Peng moved to New York City in 1973 and some believe he brought General Tso's chicken with him (Bateman 2016). It is also argued that chefs from New York visited Taiwan and brought Peng's dish with them prior to his move to the U.S.	Following the 1882 Chinese Exclusion Act, "restaurant work was one of the few jobs that the Chinese could find, and Chinese restaurant owners discovered that if they adapted simple dishes to American taste, that they could make money" (source: The Search for General Tso). According to the film, chefs began to change menus based on demand, and ultimately adapted their food to appeal to white audiences By 1900, the popular dish "chop suey" had been created, combining "Americanized meats and 'exotic' flavorless vegetables," and a mixed menu of American food and Chinese American food became national phenomenon (source: The Search for General Tso). In 1940, David Leong moved to Springfield, Missouri and created cashew chicken to appeal to the local population, frying the chicken, knowing that American's in the south loved fried chicken. Fortune cookies were also invented in the US in the 1940s, unique to American Chinese cuisine. Different regions adapted Chinese food to fit the local demand. In Louisiana, they serve Chinese gumbo with alligator meat. General Tso's Chicken became an American phenomenon in the 1970s when Michael Tong and chef T. T. Wang opened the Shun Lee Palace in NY. Wang takes credit for inventing the dish, but he was actually inspired by Peng Chang-kuei's invention in Taiwan, and added more sugar because he thought the American palate was sweeter than the Chinese (source: The Search for General Tso).

1.2 Links to Traditional Scientific Research

We document the extent of herbal ingredients in the scientific literature. Using our list of herbs, we search for scientific articles listed on PubMed containing any of our herbs as dietary supplements. Towards this, we utilized the PubMed Dietary Supplements Subset. The PubMed Dietary Supplements Subset allows researchers to retrieve dietary supplement related citations on topics such as traditional Chinese medicine and herbal medicine, among other topics. Queries of herbal ingredients will return dietary supplement related articles on PubMed that contain the queried ingredients. Searches using our herbs resulted in 658,488 articles on PubMed, published in 11,974 unique scientific journals.

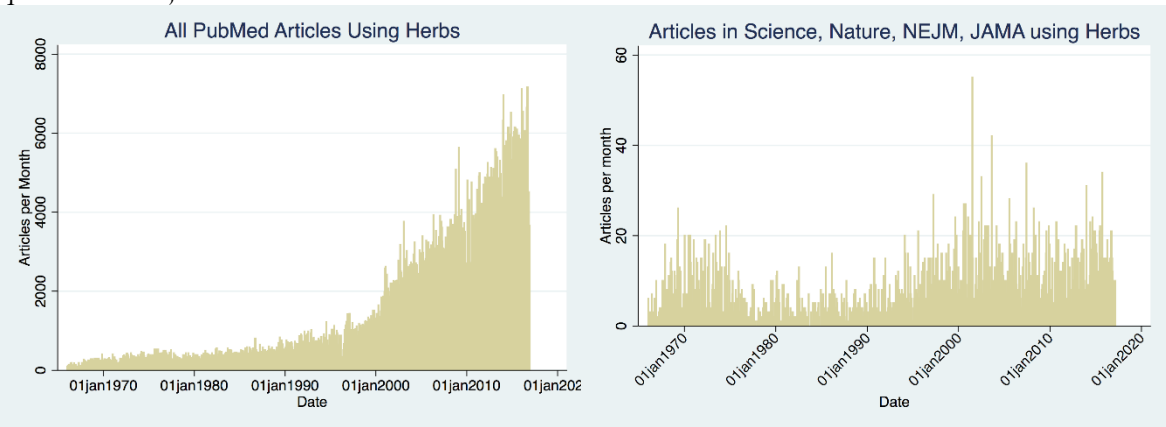


Figure A1. *Articles in scientific research using herbal supplements* Frequency plots of scientific articles on PubMed over time. We see a general increase in articles using herbal supplements over time. Restricting the subset to the most impactful journals also show a general increase in herbal research.

Table A2. *NBO Company Shares of Herbal/Traditional Products: % Value 2012-2016*

% retail value rsp	2012	2013	2014	2015	2016
Mondelez International Inc	7.8	7.8	7.7	7.4	6.9
Procter & Gamble Co, The	3.4	3.0	3.1	3.0	2.9
Ricola Inc	2.7	2.9	2.9	3.0	2.9
GSK Consumer Healthcare	-	-	-	1.3	1.5
Prestige Brands Inc	1.3	1.3	1.3	1.3	1.3
McNeil Consumer & Specialty Pharmaceuticals	1.3	1.2	1.1	1.0	1.0
NBTY Inc	1.0	1.0	0.9	0.9	0.8
NFI Consumer Products	0.2	0.3	0.5	0.7	0.8
Herbalife International Inc	0.8	0.9	0.8	0.8	0.8
General Nutrition Centers Inc	0.8	0.8	0.8	0.8	0.7
Forever Living Products LLC	0.8	0.8	0.8	0.7	0.7
Korea Ginseng Corp	0.4	0.4	0.5	0.6	0.7
Haw Par Healthcare Ltd	0.5	0.5	0.5	0.6	0.6
CNS Inc	0.9	0.7	0.7	0.6	0.6
Amway Corp	0.6	0.6	0.6	0.6	0.6
Nature's Way Products Inc	0.4	0.5	0.5	0.5	0.5
Performance Health Inc	0.4	0.4	0.5	0.5	0.5
Chattem Inc	0.6	0.5	0.5	0.4	0.4
Perfecta Products Inc	0.2	0.3	0.3	0.4	0.4
Nutraceutical International Corp	0.3	0.3	0.3	0.4	0.4
Wakunaga Pharmaceutical Co Ltd	0.4	0.4	0.4	0.4	0.3
Lily of the Desert Organic Aloeceuticals	0.3	0.3	0.3	0.3	0.3
Nature's Sunshine Products Inc	0.3	0.3	0.3	0.3	0.3
Troy Healthcare LLC	0.3	0.3	0.2	0.2	0.2
Pfizer Consumer Healthcare Inc	0.3	0.2	0.2	0.2	0.2
Concepts in Health	0.5	0.4	0.3	0.3	0.2
Windmill Health Products	0.2	0.2	0.2	0.2	0.2
DSE Healthcare Solutions LLC	-	0.2	0.2	0.2	0.2
Alan James Group LLC	0.2	0.2	0.2	0.2	0.1
Smith Bros Co, The	0.1	0.1	0.1	0.1	0.1
Novartis Corp	0.7	0.5	1.0	-	-
WF Young Inc	0.2	-	-	-	-
Other Private Label	0.6	0.6	0.6	0.5	0.5
Others	71.5	72.0	71.5	71.8	72.4
Total	100.0	100.0	100.0	100.0	100.0

Source: Euromonitor International from official statistics, trade associations, trade press, company research, store checks, trade interviews, trade sources

Section 2. Validity of the visa shock

2.1 Labor Condition Applications

One concern is that the H-1B visas were targeted towards large IT companies, and firms writing herbal patents would not have been affected by the visa shock. While we are not able to observe the number of H-1B visa grants at the firm level, we can observe the number of Labor Condition Applications (LCAs), a prerequisite of H-1B visas, applied for by each firm. In this section, we argue that herbal patent assignees benefit more from H-1B visas than the average firm does. Back of the envelope calculations show that the visa shock allowed them to hire 424 Chinese/Indians instead of the 141 under stricter immigration laws, and led to 68 additional patents.

These numbers are consistent with the patenting and hiring rates of Amgen, suggesting our calculations are plausible.

The Foreign Labor Certification Data Center provides historical data on LCAs issued since 2000. We take all 633 herbal patent assignees located in the US that filed for herbal patents, and match them to a list of entities that filed LCAs between 2000 and 2016. This includes all U.S. based assignees active during the visa shock period who have filed a herbal patent; in some cases the herbal patents filed may not have been granted. In fact, the sample size drops to 401 assignees if we consider only assignees who have been granted a herbal patent. Using a fuzzy string matching algorithm, we calculate all pairwise string similarity scores between patent assignees and LCA filing entities, and manually inspect matches to create a cutoff score above which we will consider entities to be a match. (In the following discussion, we use a score of 93 as the cutoff, but the results are similar to using cutoffs of 94 or 95).

We next show herbal patent assignees file more LCAs than other firms on average, and thus are likely to have more H1B hires than the average firm. We use the matched sample from above and plot the quantile-quantile plot of total LCAs filed by our assignees and all other firms (Figure A2). The quantile-quantile plot does not follow the 45 degree line, implying the two distributions are different. Furthermore, we see the number of LCAs filed by herbal patent assignees is left skewed, suggesting that herbal assignees are more likely to hire more people through the H1B visas than other firms filing LCAs. T-test results show that herbal assignees file for 146.7 more LCAs (*t*-statistic 4.81) than other firms, further showing herbal patent assignees are a major beneficiary of the H-1B visas.

Going forth, we make three assumptions that will allow us to measure the relationship between hiring ethnic migrants and patenting. First, we assume LCA filings for treated assignees are directly correlated with their H1B grants, and because 0.1666% of LCA filings are by treated assignees, they will collect 0.1666% of the H1B visas. Second, we assume that 40% of H1B visa grants during 1999-2003 go to inventors with Chinese/Indian nationalities. Our third assumption is that we can estimate the number of patents granted to treated firms during the shock period by predicting counterfactuals using our difference in difference specification.

Our calculations suggest the tripling of the visa cap allowed firms to hire 283 more Chinese/Indian inventors, in addition to the 141 they would have hired under stricter immigration laws. We arrive at this conclusion as follows. The total number of LCAs filed between 2001 and 2016 is 25,128,680, and the number of LCAs filed by capped herbal patent assignees is 41,874. This shows that capped herbal patent assignees file 0.1666% of all LCAs. If capped herbal patent assignees secured H-1B visas in the same proportion as the filed LCAs, this suggests that of the 636,994 H-1B visas issued during the visa shock period, 1999-2003, our firms would have used around 1,061. If the fraction of Chinese/Indian inventors is around 40%[28] (an assumption borrowed from a report filed by the Center of Immigration Studies or CIS), herbal patent assignees would have hired 424 new Chinese/Indian inventors[29]. Roughly two thirds, or 283 inventors, would not have been available for hire under the stricter visa regulations.

We next show around 17.45 percent of these inventors are granted patents, and this leads to roughly 68 new patents (or 0.1604 patent per Chinese/Indian H1B hire). During the visa shock period, we observe 74 new Chinese/Indian inventor names in our patent data, suggesting about 17.45 percent of new hires file and are granted patents (we get this by dividing the 74 new

[28] CIS documents show that in 2005, the fraction of Chinese/Indian H-1B beneficiaries was less than 45%: https://www.cis.org/Report/Wages-H1B-Computer-Programmers

[29] We only observe new inventors for granted patents, and around 63% of assignees that file herbal patents are granted one.

Chinese/Indian names that we observe in the patent records, by the total of 424 new Chinese/Indian inventors that we estimate were hired under H-1B at the capped firms). We further calculate a rough estimate of the number of additional herbal patents filed by predicting patent counts with our fitted regression models. First, we obtain predicted patent counts by plugging in our data into the model with time and assignee fixed effects (column 4 of Table 2) and exponentiating the results. After obtaining the predicted patent counts, we predict a counterfactual log patent count by setting the coefficient on the interaction to zero, again plugging in our data and exponentiating, and take the difference of the two predictions. Summing up the differences, we see that the visa shock generated an additional 68.06 patents, or 0.1605 patents per Chinese/Indian H1B hire.

Finally, we show that the estimates above align with a well-known pharmaceutical firm's H1B hiring and patenting activities. We are able to observe H1B hires and patenting for Amgen in 2015[30]. In 2015, they filed for 420 LCAs, 80 of which led to H1B visas. In addition, because 80% (note that the percentage of H1Bs granted to Chinese/Indians increases in 2015 compared to the 2005 fraction according to the USCIS reports) of H1Bs were granted to Chinese/Indians in 2015, around 64 of these hires would be Chinese/Indian nationals. We calculate the number of herbal patents by Chinese/Indians by counting the number of herbal patents by Amgen in 2015 in our data. We observe Amgen filed 71 herbal patents that year, 9 of which were granted. This suggests that 1 additional Chinese/Indian inventor amounts to 0.1406 herbal patent grants, an estimate similar to the 0.1605 obtained above.

Thus, our back of the envelope calculations suggests that an increase in the visa cap led to 424 new hires of Chinese/Indian ethnicity by herbal patent assignees, 74 of which (around 17.45 percent) were granted a patent during the same period. Hiring these inventors led to an increase of 68 herbal patents for our assignees. Hiring patterns and patenting patterns of a large firm are consistent with our calculations, adding to the credibility of our methods.

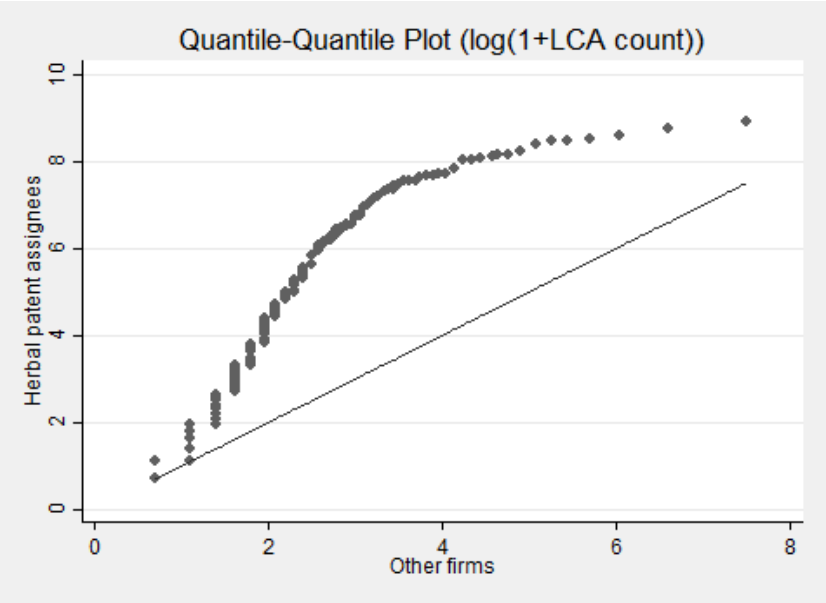


Figure A2. Quantile-Quantile plot of herbal patent assignees and all other organizations that filed for an LCA

[30]https://www.uscis.gov/sites/default/files/USCIS/Resources/Reports%20and%20Studies/Immigration%20Forms%20Data/BAHA/h-1b-2015-employers.pdf

2.2 Visa shock increased Chinese/Indian inventor hiring

We further test whether the visa shock increased the likelihood of hiring new Chinese/Indian inventors. Towards this, we estimate the following regression equation

$$1(HiredEthnic)_{jt} = \alpha + \beta_1 Capped_j + \beta_2 Shock_t + \gamma Capped_j \times Shock_t + \phi_j + \lambda_t + \varepsilon_{jt}$$

Our main dependent variable is an indicator variable for whether firm j hired any Chinese/Indian inventors at time t. We present the results in Table A3.

Table A3. *Effect of visa shock on hiring Chinese/Indian inventors*

	(1)	(2)	(3)	(4)
	Dep Var: 1(Hired Chinese/Indian inventor)			
Capped x Shock	0.02	0.02	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Capped	-0.00	0.00	-0.00	
	(0.00)	(0.00)	(0.00)	
Shock	0.01	0.01		
	(0.01)	(0.01)		
Constant	0.01	-0.00	-0.02	-0.06
	(0.00)	(0.00)	(0.00)	(0.00)
Controls	N	Y	Y	Y
Time FE	N	N	Y	Y
Firm FE	N	N	N	Y
Observations	8998	8998	8998	8998
Adjusted R^2	0.005	0.039	0.041	0.021

Note: Standard errors in parentheses, clustered at the assignee level. Observations at the assignee-year level, for all years an assignee (firm or University) was in operation. We use the tsfill command in Stata to fill in missing assignee-year pairs. Dependent variable is an indicator for whether the firm hires a Chinese or Indian inventor in a year. Controls include the fraction of Chinese/Indian inventors, firm age, inventor count, and total number of Chinese and Indian inventors. *p*-values for the interaction term (row 1) across columns are 0.030, 0.045, 0.027, and 0.028 respectively.

In the baseline specification, we see that firms subject to the visa cap increased hiring of Chinese/Indian inventors by 1.81 percent, compared to exempt firms. The p-value for this coefficient is 0.03, again suggesting that these effect sizes are highly unlikely under the null hypothesis of no effect. Compared to the baseline likelihood 1.06 percent of hiring a Chinese/Indian inventor, we see the visa shock increased the likelihood of hiring a Chinese/Indian inventor by 171 percent. The effect size is similar after including controls (Column 2), time fixed effects (Column 3), and firm fixed effects (Column 4). The p-value for our final result is 0.028, and the coefficient suggests that treatment increased the likelihood of hiring a Chinese/Indian inventor by 187 percent compared to the baseline. Interestingly, we do not see large differences in the time invariant likelihood of hiring Chinese/Indian inventors across capped and exempt firms, nor in the time effect (second and third rows respectively).

We repeat the lead-lag analysis as above, and plot the resulting coefficients in Figure A3.

Again, the 95% confidence intervals always include zero before the shock, suggesting that there are no pre-trends. Here we see the effect of the visa shock on hiring over time as well. While in 1999 and 2000 the confidence intervals do not include zero, from 2001 onwards we fail to reject a null effect. This echoes the fact that the visa quota was met in 1999 and 2000, but not in 2001-2003[31].

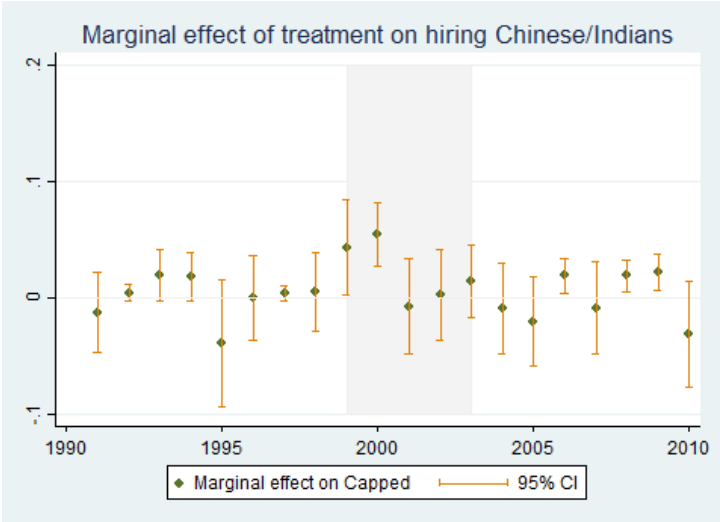


Figure A3. *Leads and lags for DD specification for hiring Chinese/Indians*

Section 3. Ethnic inventors bring in new knowledge

In this section, we test whether new inventors, and in particular new ethnic inventors, are more likely to introduce knowledge that was previously geographically locked. One implication of knowledge that is previously locked is that it would be less familiar in the US context, in our case measured by whether a patent's herbs are new, and how familiar the herb is (as measured by Google Ngrams). Then, a testable hypothesis is whether new inventors, in particular new Chinese/Indian inventors, are more likely to introduce herbs that are less familiar in the US context. We test this hypothesis by estimating the following regression equation using ordinary least squares:

$$\boldsymbol{New\ Herb_{ijt} = \beta_0 + \beta_1 New\ Inventor_{ijt} + \beta_2 New\ Ethnic\ Inventor_{ijt} + \phi_j + \lambda_t + \varepsilon_{ijt}}$$

We use two measurements for our $New\ Herb_{ijt}$ variable. In the first case, it is an indicator for whether patent i introduces a new herb at time t. In the second case, it measures the log frequency (as measured by Google N-grams) of the newly introduced herb used in patent i, applied for by firm j at time t. $New\ Inventor_{ijt}$ is an indicator for whether patent i includes a first-time inventor, and $New\ Ethnic\ Inventor_{ijt}$ is an indicator for whether patent i includes a first-time inventor who also has a Chinese or Indian name. We include firm fixed effects ϕ_j, and time fixed effects λ_t to control for firm level unobservables that may affect the types of herbs being used on a patent. The coefficient β_1 captures the change in probability of introducing an herb (and its familiarity) associated with the inclusion of a new inventor, and the coefficient β_2 measures a similar

[31] https://redbus2us.com/h1b-visa-cap-reach-dates-history-graphs-uscis-data/

change when the new inventor is a Chinese or Indian. In particular, β_2 captures the incremental effect of a new inventor being of Chinese/Indian descent on the familiarity of the herb.

We present our results of estimating equation (6) in Table A4. In columns (1-2), our dependent variable is the likelihood of introducing a new herb. Compared to the baseline probability of introducing a new herb (17.41 percent), we see that patents with new inventors are associated with a 15 percentage point increase in the probability of introducing a new patent (column 1, t-statistic = 3.57). Furthermore, in Column 2, an addition of a new Chinese/Indian inventor is associated with another 15.3 percentage point increase in the likelihood of introducing a new herb (p-value 0.041). Columns 3-4 show whether the herbs being introduced by these inventors is more or less familiar. In our sample, herbs have an average log frequency of -16.63 with a standard deviation of 2.369. Again, we see that herbs used by new Chinese/Indian inventors are 2.28 log frequencies less similar in addition to the 2.24 by new inventors, almost two standard deviations less familiar than the average herb. An equivalent change in terms of herb names would be from basil (-11.897) to poria (a type of Chinese mushroom, -16.635), or from buckwheat (-14.257) to Picrorhiza (a Himalayan herb, -19.070). While this relationship is by no means causal, there is indeed a strong association between new inventors, especially new ethnic inventors, with filing new and unfamiliar herbs.

Table A4. *New Chinese/Indian inventors use less familiar herbs*

	(1)	(2)	(3)	(4)
Dep Var:	Pr(Introduce new herb)		Log frequency of new herb	
New inventor	0.150	0.122	-2.666	-2.244
	(0.042)	(0.044)	(0.737)	(0.776)
New ethnic inventor		0.153		-2.286
		(0.075)		(1.147)
Controls				
Number of claims	0.004	0.004	-0.051	-0.051
	(0.002)	(0.002)	(0.029)	(0.029)
Has Chinese/Indian	0.037	-0.041	-0.720	0.455
	(0.120)	(0.111)	(1.985)	(1.824)
Inventor count	0.027	0.028	-0.263	-0.274
	(0.012)	(0.011)	(0.158)	(0.147)
Chinese/Indian count	-0.030	-0.050	0.479	0.785
	(0.078)	(0.077)	(1.277)	(1.278)
Number of herbs	0.000	0.000	0.014	0.013
	(0.003)	(0.003)	(0.026)	(0.026)
Constant	0.023	0.151	-1.255	-3.174
	(0.095)	(0.116)	(1.701)	(1.999)
Assignee FE	Y	Y	Y	Y

Year FE	Y	Y	Y	Y
Observations	758	758	758	758
Adjusted R^2	0.162	0.170	0.148	0.155

Note: Results of estimating equation (6) using OLS. Standard errors in parentheses, clustered at the assignee level.

Section 4. Firm implications

We next provide suggestive evidence regarding the value of herbal patents by first generation ethnic migrant inventors (as measured by citations). In our context, the value will depend on the quality of the knowledge that was previously locked in the ethnic migrant inventor's home region, and on the host country's firms' ability to codify this knowledge. Thus, given similar access to ethnic migrant inventors, firms that are quicker to codify the new knowledge ethnic migrants carry will benefit more. Ideally, we would test this by randomizing firms' abilities to extract and codify new knowledge by ethnic migrants. Since we are not able to do so, we provide correlational evidence that among firms hiring new ethnic inventors (capped firms), firms that are quicker to patent new herbal knowledge will accrue more citations. We test this by comparing the number of citations to patents with new herbs by capped firms to those without.

Below, we present a table showing that patents filed by capped firms during the visa shock period accrue more citations if they include new herbs. We see that patents with new herbs filed by capped firms during the shock period had more citations than any other group. This suggests that when firms have expanded access to ethnic migrant inventors (capped firms during the visa shock period), they are able to generate more valuable patents by introducing new herbs.

Table A5. Average citations per patents with new herbs and without

	Patent has no new herb		Patent has new herb	
	Non-shock period	Shock period	Non-shock period	Shock period
Exempt	8.89	12.05	5.67	7.00
Capped	14.27	15.98	8.99	16.31

Note: Shock period denotes years between 1999-2003.

We formalize this notion, and estimate the following regression using ordinary least squares:

$$\log(1 + Citation\ Count)_{ijt} = \delta Capped_j \times Shock_t \times NewHerb_i + \beta_1 NewHerb_i + \gamma_1 Capped_j \times Shock_t + \gamma_2 Capped_j \times NewHerb_i + \gamma_3 Shock_t \times NewHerb_i + \lambda_t + \phi_j + \varepsilon_{ijt}$$

The coefficient of interest is δ, which denotes the marginal percent increase in citations when a new herb is included on an herbal patent, when the visa cap is relaxed. We include all interactions between our three indicator variables, *Capped*, *Shock*, and *NewHerb*. We control for year fixed effects and assignee fixed effects, and hence *Capped* and *Shock* are dropped. We present estimation results below.

We see that having a new herb on a patent is correlated with a 91 percent increase in citation counts. This suggests that within patents filed by capped firms (who could hire more ethnic migrant inventors), the ones that introduced new herbs turned out to be more valuable. We are careful not to attach a causal interpretation to this measure. For instance, inventor ability may be driving both citation counts and using new herbs, and our results are reflecting better screening abilities of firms.

Table A6. *Effect of visa shock on herbal patent citations*

	(1)	(2)	(3)

Model:	OLS		Poisson
Dependent Variable:	Log(1+citations)	Log(1+Citations)	Citations
Capped x Shock x NewHerb	0.617	0.651	0.638
	(0.379)	(0.365)	(0.338)
NewHerb x Post	-0.166	-0.190	-0.039
	(0.310)	(0.301)	(0.208)
NewHerb x Capped	-0.513	-0.495	-0.597
	(0.200)	(0.187)	(0.346)
Capped x Post	-0.313	-0.349	-1.035
	(0.334)	(0.336)	(0.734)
Constant	1.884	1.782	
	(0.293)	(0.304)	
Year FE	Y	Y	Y
Assignee FE	Y	Y	Y
Controls	N	Y	Y
Observations	758	758	497
Adjusted R^2	0.127	0.126	-

Standard errors in parentheses, clustered at the assignee level.

Section 5. Alternate specifications

In this section, we present further results using our difference in difference setting. First, we show that our results are robust to relaxing the assumptions we made regarding the firm founding dates, and allowing firms to have delays between firm founding dates and the initial patent application dates. Second, we show that our results are robust to using nonlinear count models, specifically negative binomial and Poisson models. Third, we show that the visa shock has a causal impact on introducing new Chinese/Indian inventors to capped firms. Finally, we show that patents with new Chinese/Indian inventors are more likely to introduce new herbs that are unfamiliar in the US context.

5.1 Relaxing assumptions regarding firm founding year

In our baseline results, we made an assumption on the relationship between the first time an assignee files a patent, and its founding date. Specifically, we try to collect founding dates from Capital IQ for our assignees. When this method fails, we use the earliest application date for that assignee, for all granted USPTO patents as the assignee's founding year. This assumes that a firm is founded upon patent filing. However, in general, the initial patent application date and firm founding date are likely to be different. Most startups do not have patents upon founding, and in the case of very old companies (founded before 1977, when the patent data starts), there may be significant lags. This section presents results relaxing this assumption, and re-validating our results.

We proceed by varying the assumption on the time it takes from the organization founding

date to filing the first patent. We first experiment with a 3 year difference, which can be a reasonable upper bound for most start-ups. We also experiment with 6, 9, and 12 years, in case we have older private firms that are not captured by Capital IQ or the USPTO database. We do not alter the founding dates of firms that we are able to obtain data from Capital IQ. We are only adjusting the founding dates of those assignees that we must impute the founding date from the initial patent application date.

Table A7. *Relaxing firm founding date assumption and linear models*

	(1)	(2)	(3)	(4)
	Dependent Variable: Log patent count (OLS)			
Time to patent:	3 years	6 years	9 years	12 years
Capped x Shock	0.040	0.041	0.042	0.042
	(0.020)	(0.016)	(0.014)	(0.013)
Time FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Observations	9239	9475	9698	9911
Adjusted R^2	0.068	0.068	0.070	0.072

Standard errors in parentheses, clustered at the firm level.

We see from Table A7 that regardless of the lags between firm founding and patent application date, we see similar results. Across all columns, the coefficients do not vary significantly. In all specifications, we control for Assignee fixed effects, time fixed effects, and other controls. In this section, we present results using the dataset with significant lags.

5.2 Nonlinear results

Our outcome variable is the number of patent counts at the firm-year level. Since this is a count variable, its values are always nonnegative, and have discrete differences. In these situations, some assumptions of OLS may not be met. The literature tends to either take logs of the count variables, or use nonlinear estimation models to overcome such drawbacks. In our baseline, we opted for the first method and took the log number of patent counts plus one, for two benefits. First, the interpretation of coefficients is more intuitive. Second, nonlinear models in difference-in-difference settings are complicated, and while they produce coefficients that have the same sign as the true difference in difference effect, more generally the signs do not need to be consistent for interaction terms.

In this section, we present results using nonlinear count models. We focus on two models: the fixed-effect negative binomial model, and the quasi-maximum likelihood (QML) estimates based on the fixed-effect Poisson model. Our data is over-dispersed (i.e., variance is greater than the mean), and thus is more suited to estimation through the negative binomial model. We do not use a zero-inflated model because our data does not suffer from zero-inflation, as discussed in the text[32].

[32] We find no evidence for an excessive number of zeros. We follow Cameron and Trivedi (2010) and compare the predicted probabilities of zeros in a Poisson distribution to the observed probabilities. We see a slightly higher

The alternative method has the benefit that even if the underlying model is incorrectly specified, the standard errors are consistent. Furthermore, QML standard errors are robust to arbitrary serial correlation patterns, and are robust to concerns of underestimated standard errors common in difference-in-difference settings. We present our results from estimating the two models below.

Table A8. *Nonlinear model results*

Dependent Variable: Patent Counts	(1) Fixed Effects Negative Binomial	(2) Fixed Effects Poisson (QML s.e.)
Capped x Shock	0.477	0.445
	(0.230)	(0.264)
Constant	-2.517	
	(1.027)	
Time FE	Y	Y
Assignee FE	Y	Y
Controls	N	Y
Observations	9911	9911
Log likelihood	-1641.332	-1664.330

Standard errors in parentheses. We relax assumptions on firm founding dates as in 5.1. Column (1) uses the negative binomial model to estimate coefficients, and column (2) uses a Poisson model with quasi-maximum likelihood estimates of the standard errors.

We see that the number of patents increased during the visa shock period. We do not include controls in the specification for column (1) because of convergence issues. For all models, the coefficient is positive, suggesting a positive effect of the visa shock on patenting. For columns (1), the effect size is significant with p=0.038, and for column (2), p=0.092.

5.3 Triple differences

In a previous version of this paper, we estimated a triple differences model for the impact of the visa shock on the likelihood of observing a Chinese/Indian inventor. First, for each herbal patent, we obtained a non-herbal (control) patent that has the same clinical use as the focal herbal patent. We measure whether we are more likely to see Chinese/Indian inventors on herbal patents than on non-herbal patents. This matching captures the notion that Chinese/Indian inventors are more likely to bring knowledge about herbs to the US. We found that it is more likely that Chinese/Indian inventors file herbal patents non-herbal patents that have similar clinical use.

We incorporate our difference-in-difference setting to this result. In the previous paragraph, we hinted that Chinese/Indian inventors file herbal patents at a greater rate than non-

probability of zeros in our observations than the Poisson model would predict (0.0089), but we see no difference in predicted counts of zeros using Stata's countfit command, and the contribution of zeros to the Pearson Chi-Square statistic is 0.001, further showing our data does not suffer from over-inflation of zeros.

Chinese/Indian inventors. The difference-in-difference setting allows us to see whether this effect is being driven by first generation Chinese/Indian inventors. Assuming the visa shock increased the number of first generation migrant ethnic inventors, we see whether the increase in ethnic inventors increases the rate at which herbal knowledge is codified. This allows us to specify a triple difference model, with all pairwise interactions between 1) control and herbal patents, 2) capped assignees and exempt assignees, and 3) during the visa shock and not. The triple interaction term here measures the increase in the rate of codification of herbal knowledge, caused by first generation migrant inventors.

We present the triple-difference results below.

Table A9. *Triple Difference Estimates*

	(1) Fraction Ethnic	(2) Has Ethnic	(3) Fraction European	(4) Has European
Herbal Patent (HP)	0.0948**	0.165**	-0.0934**	-0.00674
	(0.0442)	(0.0816)	(0.0461)	(0.0340)
Cap-subject assignee (CS)	0.00545	-0.0356	-0.0865***	-0.0955***
	(0.0224)	(0.0420)	(0.0280)	(0.0225)
Shock	0.00190	0.0264	-0.0820**	-0.0400
	(0.0303)	(0.0639)	(0.0412)	(0.0348)
Herbal x Shock	-0.0410	-0.145	0.0894	-0.0431
	(0.0619)	(0.107)	(0.0696)	(0.0677)
Capped x Shock	-0.0133	-0.0431	0.0972**	0.0508
	(0.0327)	(0.0664)	(0.0444)	(0.0381)
Herbal x Capped	-0.0746	-0.145*	0.0346	-0.0299
	(0.0461)	(0.0836)	(0.0495)	(0.0379)
DDD	0.116*	0.247**	-0.146*	0.0117
	(0.0665)	(0.113)	(0.0747)	(0.0715)
Time Trend	0.00533***	0.00823***	-0.000279	0.00100
	(0.000744)	(0.00117)	(0.00138)	(0.00136)
Citations Count	-0.000166	0.000176	0.00149***	0.00130***
	(0.000179)	(0.000304)	(0.000268)	(0.000217)
Inventor Count	0.0215***	0.0629***	-0.0362***	0.0106***
	(0.00570)	(0.00632)	(0.00401)	(0.00292)
Constant	-10.62***	-16.41***	1.499	-1.091
	(1.489)	(2.347)	(2.771)	(2.712)
Observations	4148	4148	4148	4148
Adjusted R^2	0.062	0.138	0.079	0.020

This table presents estimation of equation (1) using ordinary least squares. The dependent variables are the fraction of or an indicator for Chinese/Indian inventors for a given patent (columns (1)-(2)), and the fraction of or an indicator for European inventors (columns (3)-(4)). Herbal and Capped are indicators for whether a patent is an herbal patent or whether the assignee is subject to the H-1B visa cap. Shock is an indicator for when the visa cap was increased (years 2000-2005). Standard errors are clustered at the assignee (employer) level. Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We see that the fraction of herbal patents with Chinese/Indian inventors increased significantly, suggesting that the rate of codification of herbal knowledge is greater for first

generation migrant inventors. This suggests that migrant inventors enter the host country and bring knowledge that is previously not in the host context.

5.4 Permutation test

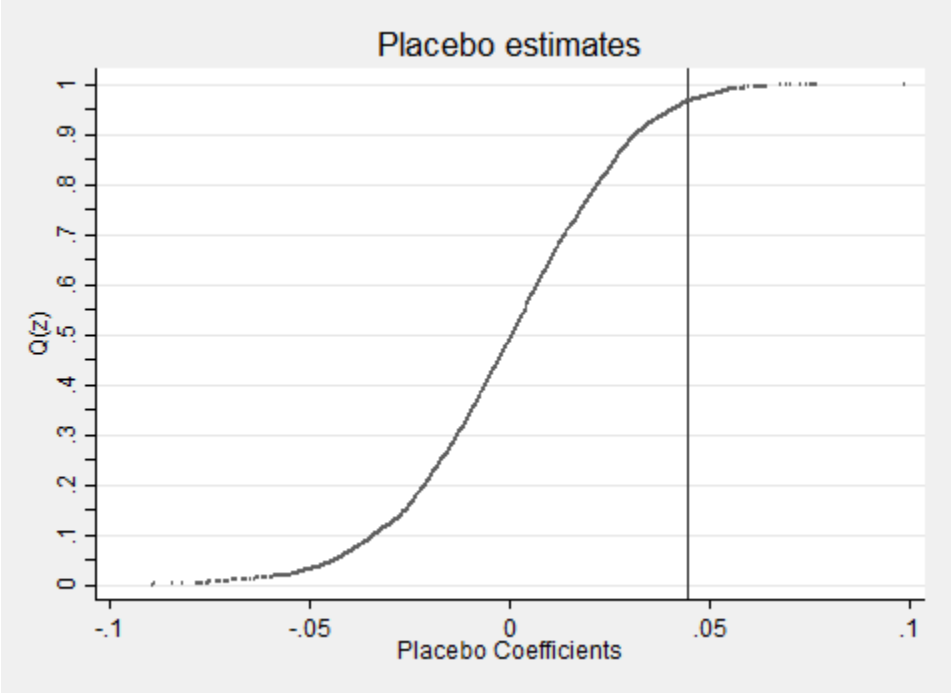


Figure A4. *Cumulative distribution of coefficients from placebo test.* From our sample of 401 herbal patent assignees, we randomly select 73 assignees to be exempt from the visa cap, and also randomly select a consecutive 5-year period to be our placebo H1B visa shock period, and run specification (1) as above, saving the coefficient on the difference in differences (DD) estimate. We repeat this process for 3,200 random treated firm-treatment window pairs. We select random placebo pairs based on two dimensions – assignment of the visa cap to assignees (done 100 times each), assignment of a visa shock period (done 32 times each for the 32 different possible 5-year time periods) for a total of 100×32 = 3,200 random placebo pairs. We plot the cumulative distribution function of the resulting DD coefficients ($Q\hat{}(\delta)$) in Figure A4. Similar to a p-value, if the visa shock positively affected herbal patenting behavior, we would expect our coefficient to be significantly larger than random (high $Q\hat{}(\delta)$), and thus appear near the upper right tail of the cumulative distribution function. We reject the null hypothesis of zero effect of the visa shock on herbal patenting if $1-Q\hat{}(\delta)>0.05$. The permutation test does not make assumptions about the error structure, and thus is robust to concerns of serial correlation.

Section 6. Summary statistics

In this section, we provide alternative summary statistics on our herbal patents, before and after the visa shock. First, we report t-test results of patent-level variables in our dataset across patents by capped assignees and patents by non-capped assignees, for all time periods. Next, we again present summary statistics on patent characteristics across capped/non-capped assignees, but for the pre-treatment period. This second table allows us to look at any systematic differences in the patents by capped assignees and non-capped assignees.

6.1 Summary statistics across the entire sample

Table A10. *t-Test Results*

	Capped firms		Exempt firms		Difference	
	mean	sd	mean	sd	b	t
Count of Chinese/Indian inventors	0.31	0.68	0.51	0.99	0.21	(2.19)
Recombined	0.31	0.46	0.29	0.46	-0.02	(-0.39)
Fraction of inventors Chinese/Indian	0.10	0.23	0.16	0.27	0.05	(2.04)
Inventors hired (All)	1.82	1.91	2.15	1.70	0.34	(1.98)
Inventors hired (Chinese/Indian)	0.22	0.59	0.35	0.89	0.13	(1.57)
Application year	2001.60	5.54	2000.17	6.57	-1.43	(-2.27)
Number of claims	15.94	13.41	14.80	10.50	-1.13	(-1.04)
Number of inventors	2.66	1.93	2.77	1.57	0.12	(0.73)
Firm age	9.74	10.12	17.70	9.43	7.96	(8.47)
Firm founding year	1991.87	9.86	1982.47	7.80	-9.39	(-11.68)
New herbs used	0.35	2.44	0.33	0.75	-0.02	(-0.15)
Herbs used	6.39	13.39	6.57	10.40	0.18	(0.17)
1(New herb used)	0.08	0.24	0.14	0.33	0.06	(1.84)
Log Ngram of least familiar herb	-16.38	2.65	-16.19	2.70	0.19	(0.71)
Log Ngram of most familiar herb	-13.47	2.71	-13.41	3.02	0.05	(0.18)
Log Ngram of median herb	-14.90	2.26	-14.74	2.46	0.16	(0.67)
Observations	635		123		758	

We see that patents by capped assignees are more likely to have, and also more likely to have more Chinese/Indian inventors. Capped firms are also more likely to hire inventors. Capped firms tend to file patents later on, but exempt firms tend to be older when they file patents. Finally, exempt firms are more likely to use new herbs.

6.2 Summary statistics before the visa shock

Table A11. *Herbal patent characteristics before visa shock (pre-1999)*

	(1) Capped		(2) Exempt		(3) Difference	
	mean	sd	mean	sd	b	t
Chinese/Indian inventor count	0.23	0.59	0.28	0.49	0.05	(0.68)
Recombined	0.25	0.44	0.25	0.43	-0.01	(-0.13)
Fraction Chinese/Indian	0.08	0.20	0.13	0.27	0.05	(1.44)
Inventors hired	1.94	1.65	2.07	1.50	0.13	(0.55)
Inventors hired (Chinese/Indian)	0.21	0.58	0.18	0.43	-0.03	(-0.44)
Application year	1995.33	4.32	1994.72	4.61	-0.61	(-0.89)
Claims	17.42	15.06	16.95	9.50	-0.47	(-0.28)
Inventor count	2.33	1.56	2.46	1.38	0.12	(0.57)
Firm age	7.34	7.67	13.53	7.53	6.19	(5.42)
Firm founding year	1987.99	8.52	1981.19	6.73	-6.80	(-6.26)
New herbs used	0.74	4.37	0.54	0.96	-0.19	(-0.56)
Herbs used	5.78	10.83	2.81	2.97	-2.98	(-3.38)
1(New herb used)	0.13	0.30	0.22	0.37	0.09	(1.62)
Log Ngram of least familiar herb	-15.81	2.25	-15.97	2.83	-0.16	(-0.38)
Log Ngram of most familiar herb	-13.14	2.34	-13.88	3.09	-0.74	(-1.64)
Log Ngram of median herb	-14.39	1.88	-14.88	2.66	-0.49	(-1.28)
Observations	189		57		246	

Both groups have a similar number of inventors and claims. Both groups' patents are equally likely to contain synthetic compounds, as measured by our *Is Synthetic* variable, and use herbs that are equally frequent in the English language. Cap exempt assignees' patents in our sample are slightly older on average, and are generally more likely to have inventors with Chinese/Indian names.

6.3 Alternate graphs using Stata's binscatter command

In this section, we plot the number of patents per assignee-year for capped and exempt firms over time. We also plot the recombination probabilities over time for any given herb, and the recombination probabilities by a specific ethnicity.

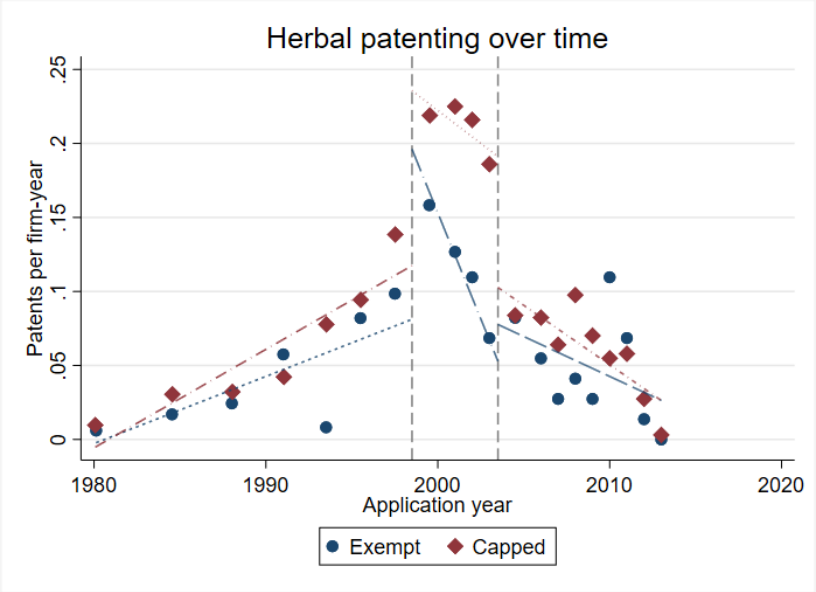


Figure A5. *Binscatter plot of patenting over time (panel A)*

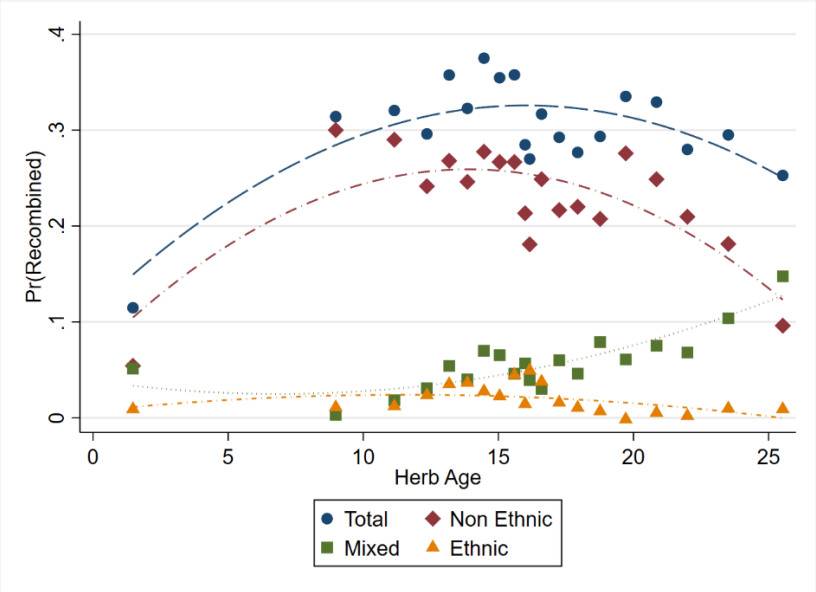


Figure A5. *Binscatter plot of recombination probabilities over time by ethnic groups (panel B)*

Section 7. Coding recombination, ethnicities, and assignees

7.1 Coding the Recombined variable

We create a variable *recombined* to indicate patent applications comprising herbs combined with synthetic compounds. To do this categorization, we used the Derwent classification for each patent. The Derwent Patent class is a manually curated standardized classification system for patents maintained by Thomson Reuters and the classification is more industry centric than technology centric. After analyzing various mixes of Derwent classes in herbal medicine patent records, we concluded that herbal medicine patent records containing Derwent classes B05, B06, or B07 comprised a mix of herbal medicines and other synthetic compounds/drugs. In the absence of any of these classes, the composition is purely made up of herbal medicines. In Derwent classification, the B class refers to 'pharmaceuticals.' Subclass B05 refers to 'other organics,' (B05 other organics -aromatics, aliphatic, organo-metallics, compounds); B06 to 'inorganics,' (inorganics -including fluorides for toothpastes etc.) and B07 to 'general'(tablets, dispensers, catheters

(excluding drainage and angioplasty), encapsulation etc.) B04 refers to 'natural products and polymers,' which also includes herbal medicine patents but does not contain synthetic compounds. B05, B06, and B07 are the only three classes in B (pharmaceuticals) that contain synthetic Western drugs. Thus, a presence of these three classes signifies a combination of synthetic compounds/drugs with herb. Fifty random abstracts of patent records having any of these three classes and 50 random abstracts of patent records with absence of all of these three classes were studied to confirm the effectiveness of using Derwent classes to code the 'Recombined' variable and this result was independently verified by two different coders and checked by the researchers.

7.2 Coding ethnicities

Probabilistically, surnames such as Xing are more likely to be associated with Chinese individuals than with other ethnicities. We build on this insight and utilized an open-source name categorizer "ethnicityguesser" to categorize inventors' ethnicities. The software is based on the Natural Language ToolKit (NLTK) package in Python, and it comes pre-packaged with a set of names and associated ethnicities. As a robustness check, we compare our ethnicity classification results when using different training sets and against Ambekar et al. (2009) who use state of the art hidden Markov models and decision trees for classification.

Table A12. *Correlations across Ethnicity measures*

	Asian1 (surname)	Asian2 (surname)	Asian1 (full)	Asian2 (full)
Chinese/Indian1 (first)	1			
Chinese/Indian2 (first)	0.9637	1		
Chinese/Indian1 (full)	0.9503	0.9195	1	
Chinese/Indian2 (full)	0.9297	0.9394	0.963	1

Note: Comparison of the two training sets provided by kitofans' ethnicityguesser. As expected, we see that for all classifications, there is a high correlation across the two measures, whether we use the full name or just the surname as the tokens.

We also compare our results to the Name Ethnicity Classifier created by Ambekar et al. (2009). If we have a high correlation between our measure of Chinese, Indian and European with the Name Ethnicity Classifier's Asian and Greater European categories, we would be confident about our measures of ethnicity. We randomly sample 10% (1,219) of our inventors' names and submit this to the Name Ethnicity Classifier's website. We present the results below. We see that 94 percent of our Chinese inventors are categorized as Asian, and 90 percent of our Indian inventors are categorized as Asian. Generally, our classification of European coincides with the categorization of Europeans by Ambekar et al (2009). Overall, the results reflect positively on our classification of ethnicities.

Table A13. *Comparison of ethnicityguesser performance to benchmark ethnicity classification product*

kitofans	Asian	GreaterAfrican	GreaterEuropean
african	5	2	0
arabic	0	0	2
chinese	**115**	**0**	**7**
czech	16	5	28
danish	1	0	25
french	12	7	170
german	1	0	54
greek	4	1	11
indian	**70**	**4**	**3**

irish	0	0	31
italian	7	2	21
japanese	133	3	2
jewish	14	11	163
korean	63	1	6
muslim	6	14	2
portugese	4	1	12
russian	0	0	3
slavic	0	0	7
spanish	7	5	51
swedish	3	1	43
swiss	2	2	36
ukranian	1	1	10
vietnamese	5	0	3

Note: Comparison of classification results using kitofans' ethnicityguesser and Ambekar et al (2009). We read the table as follows. Of the 122 names classified as Chinese using ethnicityguesser, 115 are classified as Asian in Ambekar et al (2009).

7.3 Educational background of Chinese/Indian inventors

Inventor backgrounds can also provide information about whether herbal patent inventors are more likely to be first generation migrant inventors. Our 3,182 herbal patent grants (U.S. and foreign assignees) contains 6,119 unique inventors, of which 1,208 unique inventors are of Chinese or Indian ethnicities. We randomly sample 552 inventors from the Chinese/Indian inventor population (45% of unique Chinese/Indian inventors in herbal patents sample) and attempt to search for their educational history in LinkedIn. To do so, we search for individuals in LinkedIn using the inventor's and assignee's names. If there is a profile that 1) has a match on the inventor name, 2) match for the assignee of interest 3) near the time period the patent application was submitted, we code this as a successful search. We successfully found 84 profiles on LinkedIn (15% of Chinese/Indian inventors that we looked up on LinkedIn), but we drop 20 individuals who do not list their educational details. For each Chinese/Indian inventor left, we document the educational background of the individuals. We document whether the inventor was educated solely in India, U.S., or China, or whether they were educated elsewhere and moved to the U.S. Of the sample, about one third of the individuals were educated solely in India and the U.S. each. About 20% of individuals were educated first in China, then moved to the U.S. The remaining inventors were educated just in China (9%) or educated in India, then educated in the US (3%). In summary, a disproportionate fraction of matched Chinese/Indian inventors filing herbal patents who we looked up on LinkedIn, were educated in China/India, indicating that they were indeed first generation migrant inventors.

Table A14. *Educational background of Chinese/Indian inventors*

Educational Background	Count	Percentage
India	22	34.38%
US	20	31.25%
China to US	14	21.88%
China	6	9.38%
India to US	2	3.13%

Total	64	100%

We also look at whether herbal patents have more inventors that were educated in China/India. Herbal patents are much more likely to have inventors educated solely in India, and similarly for Chinese educated individuals. On the other hand, inventors educated abroad who moved to the US are less likely to write herbal patents. Inventors educated solely in the US are less likely to write herbal patents, despite their being ethnically Indian/Chinese.

Table A15. *Educational background of Chinese/Indian inventors by patent type*

Educational Background	Control Patent	Herbal Patent
India	5	17
US	14	6
China to US	9	5
China	3	3
India to US	2	0
Total	33	31

Finally, we see whether the visa shock increased the number of foreigners writing patents. Towards this, we look at whether patents written during the visa cap increase have more inventors that were educated outside the US. The shock seems to have increased the proportion of Indian inventors, but decreased all other types of inventors.

Table A16. *Educational background of Chinese/Indian inventors over time*

Educational Background	Non-Shock	Shock
China	3 (9.09%)	3 (9.68%)
China to US	9 (27.27%)	5 (16.13%)
India	5 (15.15%)	17 (54.84%)
India to US	2 (6.06%)	0
US	14 (42.42%)	6 (19.35%)
Total	33	31

7.4 Assignee classification

We categorized assignees into three broad groups: individuals, U.S. based firms and Universities, and Foreign firms and Universities. We merge our herbal patents dataset with two external datasets for this process: USPTO's PatentsView database, and Capital IQ. PatentsView contains disambiguated assignee data, and classifies each assignee into U.S. Company or Corporation, Foreign Company or Corporation, Individuals, and so forth. Capital IQ provides researchers with corporate headquarters for a company. In our sample of U.S. assignees, there are 401 unique assignees, 73 of which are cap exempt.

In our original dataset of the universe of herbal patents, there were a total of 7,157 patents. We obtain the geographical data for 3,183 patents from PatentsView, 3,562 from Capital IQ and 412 from manual searches. We find that 2,512 of our patents are filed by individuals, 2,851 by foreign companies, and 1,794 by firms/Universities based in the US. We also obtained a list of H1B visa cap-exempt employers from a 3rd party online employment entity[33]. The online list contains 12,479 employers who have been categorized as exempt from the H1B visa cap. We matched these

[33] Source: http://www.myvisajobs.com/Search_Visa_Sponsor.aspx

employers to our list of assignees, and further searched for "university" and "college" to construct a list of assignees that are exempt from the H1B visa cap (*CAP*). Out of the 4,179 total number of unique assignees in our herbal patent sample, 158 unique assignees are exempt from the H1B visa cap. In our sample of U.S. assignees, there are 401 unique assignees, 73 of which are cap exempt.

7.5 Sample restrictions

Our sample restrictions are based on three considerations: 1) location of firm, 2) founding dates/final patenting dates, 3) patent grants. We first explain the data sources, then delineate the sample restriction process and rationale, and finally discuss the magnitudes of these changes. We use assignee locations from PatentsView when available, and also use headquarter locations provided by Capital IQ. We also use Capital IQ to obtain founding dates when available, and impute founding dates based on patent application years. We also obtain the date of the last patent filed by an assignee. Application status (granted or not) is obtained through PatentsView.

In the full dataset, there are 4,179 assignees, 1,368 of which are granted any patents. Restricting the headquarter location to US based assignees brings this down to 1,037 firms, 585 of which are granted any patents. Finally, we restrict the sample to firms founded before 2004 that continued filing patents through 2000. We thus have 633 firms found before 2004 that filed at least one patent after 2000, 401 of which are granted any patents.

The largest change is in excluding the non-US firms, but our identification strategy does not specify in which direction foreign firms' patenting should move. The restriction on founding dates and last patenting dates is necessary because patenting outside of our visa shock period may bias the results downwards. Finally, the restriction on firms with any granted patents can be relaxed to obtain similar, but noisier results. Using all 633 firms that are filing patents, we observe a 2.29 percent increase in patent filings (p=0.0998) while including assignee fixed effects, application year fixed effects, and time varying firm controls.

Section 8. Herb characteristics

For all granted herbal patents, we collect a list of herbs mentioned in the title and abstract of the patent. For each of these herbs, we collect the Google N-gram score of the herb in the patent application year. Google N-grams provides users with the raw frequency of words in all American English books digitized by Google, for a given year until 2009. We define the Familiarity of the herb to be the log of the N-gram score. Since we have the universe of herbal patents, we can also obtain the minimum year in which an herb was used in a patent. We define the Year Introduced variable as the minimum application date across all patents using a specific herb. We present a select list of herbs, and their characteristics below.

Table A17. Most frequent and least frequent herbs

Herb Name	Patents using herb	Familiarity	Year Introduced	Notes
corn	82	-10.6883	1984	Most frequent herb
soybean	71	-13.1475	1985	
soy	66	-12.1929	1994	
green tea	65	-14.0068	1996	
vegetable oil	60	-13.5257	1984	
		⋮		
rosmarinus	12	-16.8946	1995	
echinacea purpurea	5	-16.8609	1999	Median familiar herb

vitis vinifera	1	-16.8603	2008	Median familiar herb
phytolacca	3	-16.8545	1992	
		⋮		
paullinia cupana	1	-19.2911	1987	
catharanthus roseus	1	-16.6776	1986	
zygophyllaceae	1	-17.6729	1986	
matico	1	-19.8119	1986	
salicornia	1	-16.6955	1985	Least frequent herb

Note: Most and least frequent herbs, including the median familiar herbs. Patents using herb are counts of patent grants by US based assignees (corporations and universities) that were active during 1999-2003. Familiarity is measured as the log of the Ngram of the herb when the patent application was filed; here we report the mean log Ngram across all patents using an herb. Year Introduced is the first year in our sample in which the herb was used.

Section 9. Impact of restrictive immigration policies

In this section, we present results showing that restrictive immigration policies can deter recombination.

Table A18. *Effect of restrictive migration policies post 2004*

	(1)	(2)	(3)
	Dependent variable: log(1+patent count)		
Capped x Post2004	-0.032***	-0.018**	-0.036***
	(0.010)	(0.009)	(0.010)
Capped	0.037***	0.023***	
	(0.006)	(0.006)	
Post2004	-0.005		
	(0.007)		
Constant	0.040***	-0.008	0.015
	(0.004)	(0.009)	(0.010)
Time FE	N	Y	Y
Assignee FE	N	N	Y
Observations	8998	8998	8998
Adjusted R^2	0.008	0.036	0.038

Cluster robust standard errors in parentheses, clustered at the assignee level. Observations are at the assignee-year level, for all years an assignee (firm or university) was in operation. F=We use the tsfill command in Stata to fill in missing assignee-year pairs. The assignee-year level dataset is thus an unbalanced panel consisting of 8,998 observations (an average of 22.4 years of observations for 401 assignees). The dependent variable is the log of the number of herbal patents filed by an assignee in a given year. Capped is an indicator for whether the assignee is subject to the visa cap; Post2004 is an indicator for years 2004 and onwards. Percentage changes are calculated as $100 \cdot (e^{\beta} - 1)$.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Section 10. Recombination by non-ethnic teams

It is worth noting we observe non-ethnic (no ethnic inventor) teams participating in recombination as well as mixed teams. This is puzzling because we hypothesized that recombination requires knowledge of both the local context and the foreign context, suggesting recombination should be driven by mixed teams. We argue indirect spillovers from prior collaborations and prior herbal patent codification can allow non-ethnic teams to recombine. We illustrate these mechanisms with empirical evidence.

10.1 Spillovers through co-inventors

We first test the relationship between prior ethnic exposure and recombination probabilities. Specifically, we test whether the probability of recombination varies with prior exposure to ethnic inventors, and whether that rate varies across the different types of teams. We discuss the graphical results first.

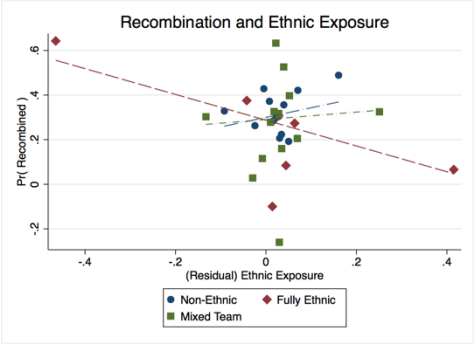


Figure A6. *Recombination probabilities and prior ethnic co-inventor exposure*

We see that for non-ethnic teams (and mixed teams), the probability of recombination increases with ethnic exposure of the patent's inventors, while fully ethnic teams exhibit a downward sloping relationship between ethnic exposure and probability of recombination. We test this formally by estimating the following equation:

$$Recombined_{ijt} = \beta_1 Exposure + \beta_2 NonEthnic + \gamma Exposure \times NonEthnic + \lambda_t + \phi_j + \epsilon_{ijt}$$

where *Exposure* is the average prior exposure to ethnic inventors across all inventors in a patent, and *NonEthnic* is an indicator variable for patents with no ethnic inventors. The main coefficient of interest is γ which captures the differential increase in likelihood of recombination from an increase in ethnic exposure for non-ethnic teams.

Table A19. *Increase in prior ethnic collaboration increases recombination*

	(1)	(2)
	Dependent Variable: Recombined	
Ethnic Exposure	-0.491*	-0.568*
	(0.260)	(0.300)
Non-Ethnic	-0.020	0.063
	(0.070)	(0.073)

Non-Ethnic x Ethnic Exposure	1.104*	1.067*
	(0.612)	(0.631)
Assignee FE	Yes	Yes
Application Year FE	No	Yes
Controls	No	Yes
Observations	501	454
Adjusted R^2	0.212	0.205

Cluster robust standard errors in parentheses, clustered at the assignee level (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Exposure is skewed, ranging from 0 to 1 with a mean of 0.028 and a standard deviation of 0.128. The sample is at the patent level, and the sample size differs because of assignee fixed effects: observations without assignee level variation will be dropped.

We see from both columns that the coefficient on "Non-Ethnic x Ethnic Exposure" is positive and significant (p=0.073 and p=0.093 for columns 1 and 2 respectively). This pattern is robust to including controls (column 2). This suggests that teams comprising non-ethnic inventors who have greater prior exposure to ethnic inventors are more likely to be apt at recombination, compared to teams of non-ethnic inventors who have lower prior exposure to ethnic inventors.

10.2 Spillovers through prior inventions

Another mechanism by which non-ethnic inventors can recombine knowledge is if the effects of the herb have been codified in the past. Codified knowledge regarding herbs significantly lowers the cost of recombination even to non-ethnic inventors. One consequence of this mechanism would be that the herbs non-ethnic inventors use will be older (and thus more likely to be codified). We perform this indirect test by estimating the following regression equation:

$$MedianHerbYear_{ijt} = \beta_1 NonEthnic_i \times Recombined_i + \beta_2 Recombined_i + \beta_3 NonEthnic_i + \lambda_t + \phi_j + \epsilon_{ijt}$$

where MedianHerbYear is the median age of the herb on a patent, NonEthnic and Recombined are indicators, and we include assignee year fixed effects and firm fixed effects when appropriate. The coefficient of interest is β_1, which compares whether recombined patents use older herbs on average, and whether this is more pronounced for non-ethnic teams.

Table A20. *Recombined patents by nonethnic inventors have older herbs*

	(1)	(2)	(3)
	Years since median herb was first used		
Non Ethnic	-2.010	-1.107	-1.526
	(1.277)	(1.071)	(1.942)
Recombined	-2.001	-1.961	-2.018
	(1.965)	(1.793)	(1.801)
Non Ethnic x Recombined	4.807	3.503	3.721

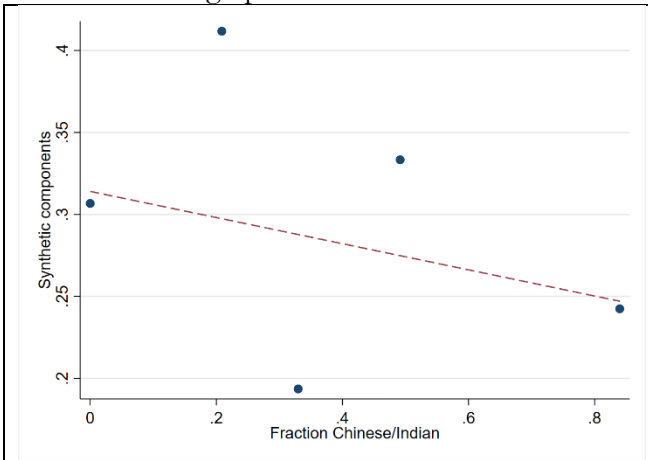
	(2.221)	(2.015)	(2.017)
Assignee FE	Y	Y	Y
Application Year FE	N	Y	Y
Controls	N	N	Y
Observations	501	494	494
Adjusted R^2	0.318	0.512	0.522

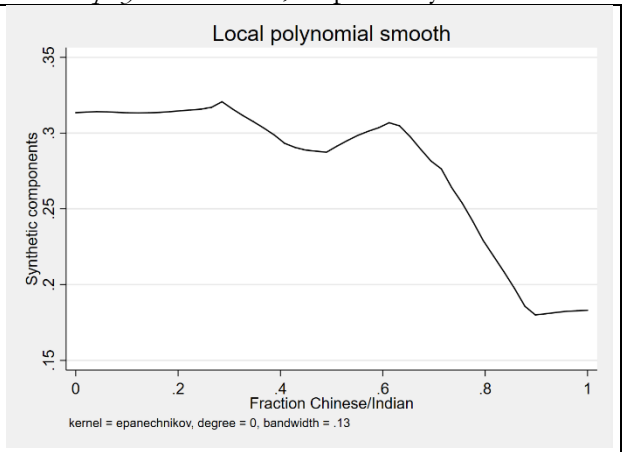
Robust standard errors in parentheses

From column 1, we see that the median herb is younger on non-ethnic teams compared to teams with ethnic inventors, but this number is statistically insignificant at the 10 percent level (p=0.116). Interestingly, the median herb in a Recombination patent is younger than reuse patents, but this effect is also statistically insignificant (p=0.309)[34]. The interaction term is positive and significant (p=0.031), suggesting that the median herb on a recombination patent is older than a reuse patent, but only for the non-ethnic teams. This is consistent with our hypothesis that non-ethnic teams need to build on prior codification of herbal knowledge, while teams with ethnic inventors do not have such restrictions.

Section 11. Alternate tests of recombination

First, to visualize the relationship, we include binned scatterplots and local polynomial smoothing estimates of the relationship between the usage of synthetic compounds and the fraction of ethnic inventors. The graphs are obtained via Stata's *binscatter* and *lpoly* commands, respectively.





[34] Note that this is only the case when controlling for assignee fixed effects. Thus, the interpretation is that the median herb in a recombination patent is younger than a reuse patent within assignees that have some ethnic inventors.

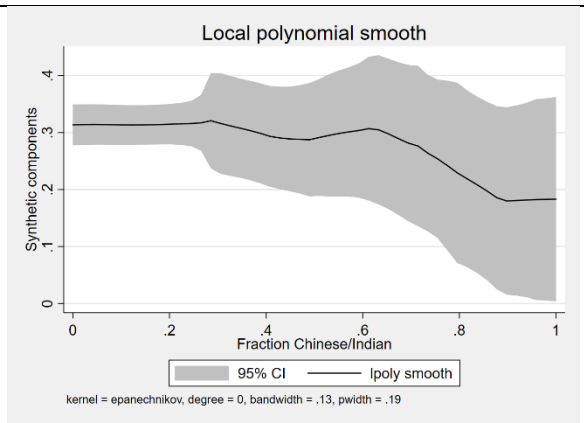


Figure A7. Binned scatterplots and local polynomial smoothing curves of recombination

The graphs suggest that there is a nonlinear relationship between the two variables. For low Chinese/Indian inventor patents, the use of synthetic compounds is relatively high, while the downward slope is primarily being driven by patents with a high fraction of Chinese/Indian inventors. We include a version of the second graph with confidence intervals as well. This last graph shows that while the probability of recombination is decreasing as the fraction of Chinese/Indians increases, the estimates are also increasingly noisier.

We next investigate whether recombination probabilities are affected by inventor counts and ethnic inventor counts independently. We estimate 3 simple models, all with recombination as the main dependent variable, and with inventor counts, ethnic inventor counts, and both inventor and ethnic inventor counts as the independent variables. We present the results below.

Table A21. *Recombination and inventor counts*

	(1)	(2)	(3)
	Dependent Variable: Recombination		
Inventor count	0.015		0.024
	(0.011)		(0.014)
Ethnic inventor count		-0.037	-0.066
		(0.032)	(0.037)
Assignee FE	Yes	Yes	Yes
Application Year FE	Yes	Yes	Yes
Observations	501	501	501
Adjusted R^2	0.216	0.214	0.218

Cluster robust standard errors in parentheses, at the patent assignee level. Sample at the patent level. Note: since we include assignee fixed effects, assignees with only one patent are dropped, hence the different sample size.

We see that while inventor count is positively associated with recombination, the effect is statistically insignificant (p=0.186). Similarly, Ethnic inventor count is negatively associated with recombination, but again statistically insignificant (p=0.251). Including both as independent variables magnifies the effect of both variables. Furthermore, the effects have greater statistical significance: p=0.097, p=0.082 for inventor count and ethnic inventor count respectively.

Section 12. Tests of inverted-U relationships

Next, we utilize Simonsohn (2017) to formally test whether an inverted U-shaped relationship exists for recombination over time. Intuitively, an inverted U-shape implies the existence of a cutoff x_c such that if $x \leq x_c$, then x and y are positively related, and if $x > x_c$, then x and y are negatively related. Simonsohn (2017) chooses the cutoff based on a "Robin Hood" algorithm (which he shows is robust to errors), estimates an interrupted regression, and checks if the coefficients for high and low values of x (1) have opposite signs and (2) are independently statistically significant according to a pre-specified threshold. We implement this test on the herb-level demeaned variables for recombination and time. That is, for patent i using herb h, we define the demeaned recombination variable $\widetilde{y_{ih}} = y_{ih} - \bar{y}$ and the demeaned time variable $\widetilde{x_{ih}} = x_{ih} - \bar{x}$ where $\bar{y} = \frac{1}{N}\sum_i y_{ih}$, $\bar{x} = \frac{1}{N}\sum_i x_{ih}$. We follow Simonsohn and estimate the cutoff $\widetilde{x_c}$ using the Robin Hood algorithm, and estimate the interrupted regression using a significance threshold of 0.05. We present graphical results below (note the y-axis denotes recombination and the x-axis denotes time since herb introduced).

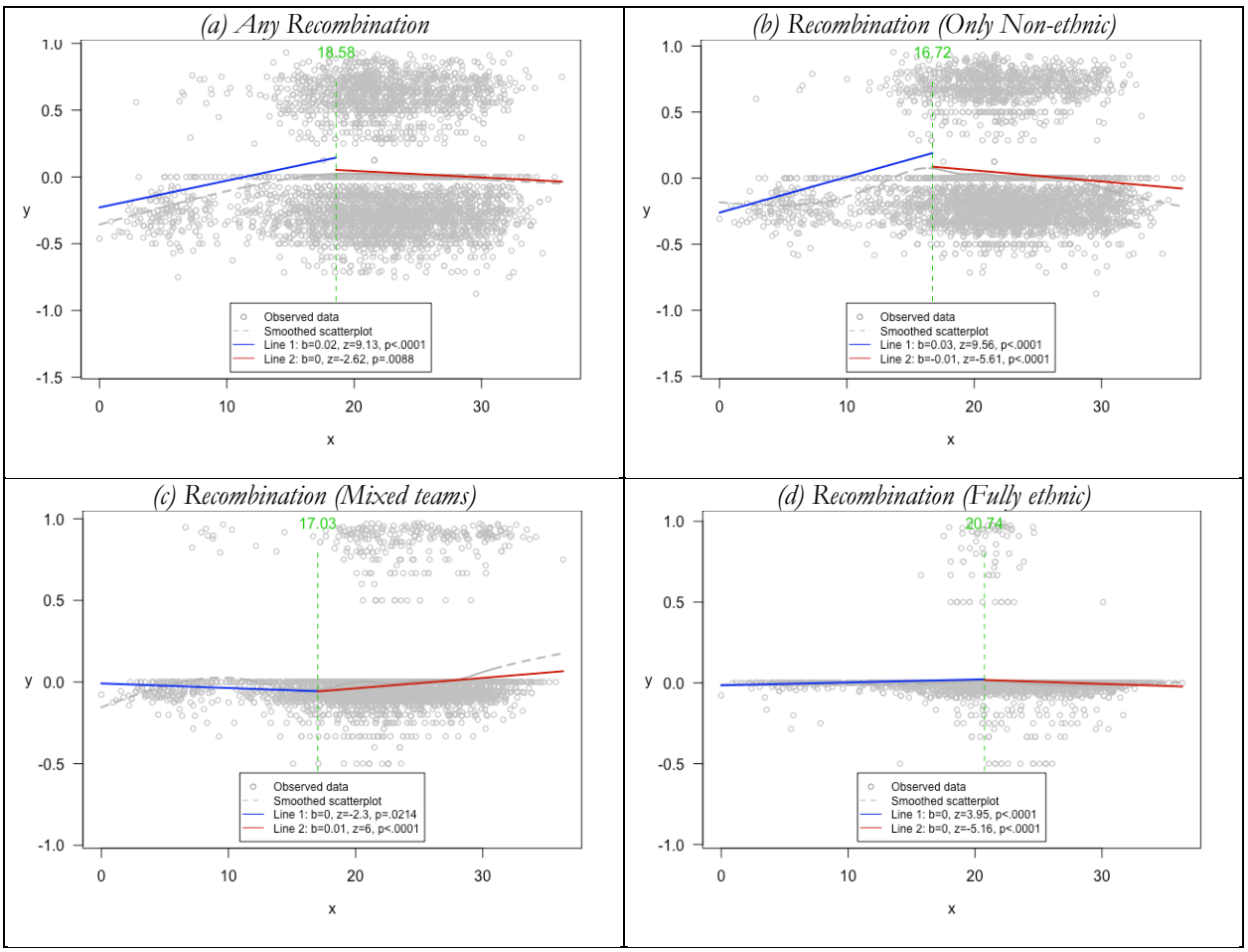


Figure A8. *Recombination probabilities over time, by ethnic teams. (Z-scores represent b/se)*

As we see from Figure A8, we observe inverted U-shape patterns for some ethnic groups (Non-ethnic and fully ethnic teams), but the opposite for mixed teams. Overall, there is an inverted U-shape relationship with recombination increasing for the first 18.6 years ($\beta_{low}^{overall} = 0.020$,

p<0.001), but decreases afterwards ($\beta_{high}^{overall} = -0.005$, p=0.009). Interestingly, recombination by entirely non-ethnic teams is increasing for the first 17 years an herb is introduced ($\beta_{low}^{nonethnic} = 0.027$, p<0.001), while recombination by mixed teams is decreasing during the same time period ($\beta_{low}^{mixed} = -0.003$, p=0.021). Subsequent recombination decreases for entirely non-ethnic teams ($\beta_{high}^{nonethnic} = -0.008$, p<0.001), while it increases for mixed teams ($\beta_{high}^{mixed} = 0.006$, p<0.001). We see similar patterns for fully ethnic teams and entirely non-ethnic teams, but the impacts are smaller ($\beta_{low}^{fullethnic} = 0.002$, p<0.001, $\beta_{low}^{fullethnic} = -0.003$, p<0.001). While not reported here, we find similar results after controlling for patent characteristics such as inventor counts and number of claims.

Qualitatively, a significant portion of initial recombination is driven by non-ethnic teams. Comparing overall recombination and non-ethnic recombination (panels a and b in Figure A8), we see the cutoff point is smaller for non-ethnic teams (16.72 versus 18.58), and the slope is larger for non-ethnic teams (0.027 versus 0.020). This suggests local inventors (non-ethnic teams) recombine knowledge faster and earlier on. The other team compositions do not significantly drive recombination during this period, with the slope for full-ethnic teams and mixed teams at 0.002 and -0.003, an order of magnitude smaller than non-ethnic teams.

Similarly, while overall recombination is decreasing as more time has passed since an herb's initial use, mixed teams continue to recombine in later time periods. Mixed teams increase the rate of recombination after 17.03 years, with recombination increasing at a rate of 0.006 per year, in contrast to the overall negative trend of recombination at -0.005.