Ethnic Migrant Inventors — Transfer and Recombination of Contextual Knowledge across Borders

Prithwiraj (Raj) Choudhury and Do Yoon Kim[1]

We study the role of ethnic migrant inventors in transferring contextual knowledge across borders and the role of ethnic inventor networks in further disseminating such knowledge. We also study microfoundations of subsequent recombination of contextual knowledge within western firms. Using a unique dataset of herbal patents filed in the U.S. by western firms and universities, we test whether contextual knowledge is codified in the west by ethnic migrant inventors and spread by their ethnic networks. Our identification comes from an exogenous shock to the quota of H1B visas, and a list of institutions that were exempted from the shock. We generate a control group of non-herbal patents that have similar medicinal purposes as our herbal patents through textual matching. Using this framework, we estimate a triple differences equation, and find that herbal patents are likely to be filed by Chinese/Indian migrant inventors and are likely to be initially cited by other Chinese/Indian inventors. We also find that Chinese/Indian migrant inventors are likely to engage in arbitraging their prior knowledge, while inventors from other ethnic backgrounds are likely to engage in knowledge recombination.

Keywords: contextual knowledge; ethnic networks; skilled migration; knowledge flows; knowledge arbitrage; knowledge recombination; microfoundations; textual analysis of patents; H1B visas

---

[1] We would like to thank Rajshree Agarwal, Gautam Ahuja, Ashish Arora, Iain Cockburn, Alfonso Gambardella, Shane Greenstein, Mariko Sakakibara, Minyuan Zhao, and seminar participants and discussants at Harvard Business School, London Business School Sumantra Ghoshal Conference, NBER Productivity Seminar and Wharton for their comments on a prior draft.
Corresponding author: Prithwiraj (Raj) Choudhury, email – pchoudhury@hbs.edu

Introduction

The innovation literature has long emphasized the importance of inventor mobility in knowledge transfer and knowledge recombination across geographic borders (Agrawal *et al*. 2006, Agrawal *et al*. 2008, Oettl and Agrawal 2008, Rosenkopf and Almeida 2003, Song *et al*. 2003). The literature starting with Jaffe *et al*. (1993) has documented that knowledge is geographically localized and as Singh and Marx (2013) have shown, though the effect of state borders within a country constraining knowledge transfers has waned with time, the effect of country borders has strengthened over time. In this literature, mobility of inventors has been suggested as a possible solution to the geographic constraints of transferring knowledge. As Rosenkopf and Almeida (2003) have articulated, mobility of inventors can serve as *bridges to distant contexts*, thus enabling firms to overcome the constraints of contextually localized search.[2] Song *et al*. (2003) showed that mobile engineers often possess technological expertise distant from that of the hiring firm and often work in non-core technological areas at their new firm. Oettl and Agrawal (2008) extend the findings of this literature and additionally document that there could also be "unintended" knowledge flows that result from the cross-border mobility of inventors, where knowledge flows accrue to firms other than the hiring firm in the new geographic region. Agrawal *et al*. (2006) and Agrawal *et al*. (2008) have also more recently shown that social proximity of inventors can serve as a substitute to physical proximity.

A relatively unexplored question in this literature relates to knowledge that is embedded in the cultural, religious and linguistic context of the home region of the inventor, and the role of inventor mobility in transferring such knowledge to a new geographic region. There is a rich literature in innovation and strategy 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 *et al*. 2010) to outline *contextual knowledge* as knowledge that is deeply embedded in its cultural, religious or linguistic context. At the time of transfer, contextual knowledge could be at the periphery of knowledge production in the host region of the mobile inventor. Given this, we study the role of the migrant inventor in transferring contextual knowledge from their home countries to western research entities in the host region. In light of the literature on the micro-foundations of knowledge recombination (Allen 1977, Fleming 2001), we also study subsequent recombination of contextual knowledge after such knowledge is transferred to western firms. Though prior literature has established a relation between inventor mobility and knowledge transfer, we know relatively less about the microfoundations of knowledge recombination after novel knowledge has been transferred by the mobile inventor to the recipient firm.

[2] This literature in innovation dates back to Porter (1990) who pointed to the emergence of dispersed geographic regions, specialized in various technologies and argued for the need of geographically distant search. Other studies relevant to the geographic localization of knowledge include Almeida and Kogut (1999) and Thompson and Fox-Keane (2005).

In this paper, we argue that ethnic migrant inventors play an important role in transferring contextual knowledge across borders. Drawing on the literature on codification of tacit knowledge (Polanyi, 1966, Dasgupta and David, 1994, Cowan and Foray, 1997), we argue that ethnic migrant inventors are deeply embedded in the context harboring the knowledge, which makes focused perception of contextual knowledge possible, understandable, and productive. When such inventors migrate to a new geographic region, they are in an ideal position to transfer contextual knowledge to the new region, using codification tools (such as patent filing) available in the new region.

We also argue that once transferred, contextual knowledge is likely to initially spread within ethnic inventor networks. The innovation literature has shown that ethnic and social ties have a strong influence on knowledge flows (Agrawal *et al*., 2006, Agrawal *et al*., 2008, Kerr, 2008, Breschi and Lissoni, 2009, Foley and Kerr, 2013). We argue that while social and ethnic proximity could determine knowledge flows for a broad set of knowledge, it is particularly relevant for contextual knowledge. We build on the innovation literature related to the impediments of transferring sticky knowledge (Von Hippel, 1994, Szulanski, 1996, Jensen and Szulanski, 2004) to argue that from the perspective of the non-ethnic inventor (i.e. inventor who does not share ethnicity and contextual proximity with the migrant inventor), contextual knowledge could initially have the property of causal ambiguity as outlined by Szulanski (1996). This could lead to an initial lack of motivation on part of the non-ethnic inventor to build on contextual knowledge even after its transfer to the new geographic region. Given this, we suggest that contextual knowledge, after its initial transfer to the new region, would initially be shared within ethnic inventor networks.

We also study the recombination of contextual knowledge, after such knowledge is transferred to the western firm. Building on the literature that looks at the microfoundations of knowledge recombination and recombinant search within firms (Allen 1977, Fleming 2001), we argue that ethnic migrant inventors are in an ideal position to arbitrage their contextual knowledge after they move to the new firm in the host region. Additionally, inventors belonging to other ethnicities, given their knowledge diversity and their familiarity with prior knowledge codified in the west, are in a better position to engage in knowledge recombination. In other words, while ethnic migrant inventors might transfer novel contextual knowledge into the boundary of the western firm, inventors of other ethnicities are likely to act as agents of ‘recombinant creation of knowledge’ (Carnabuci and Operti, 2013).

To test these propositions, we created a unique dataset of 2,060 herbal patents filed in the United States Patent Office (USPTO) between 1977 and 2010 by western firms and universities. The patent filing entities included large western multinationals such as Abbott Laboratories, Bayer Bristol-Myers Squibb, Eli Lily, Pfizer, Merck, Colgate Palmolive, Proctor and Gamble, Unilever, and large universities across the United States (U.S.). The market for products based on herbal remedies was estimated to be around $5.4 billion in 2016 and later in the paper, we provide evidence around the importance of herbal remedies to the western bio-pharmaceutical industry and more broadly to western science. Among other stylized facts we later present trends of publications related to herbal remedies in journals such as *Science*, *Nature*, and the *New England Journal of Medicine*.

We employ a novel matching technique using textual analysis to identify control patents filed in the same year and targeted at the same disease area as each herbal patent. Given this, we find that herbal patents are disproportionately likely to be filed by ethnic Chinese and Indian inventors. However, this does not help clarify whether the ethnic inventors filing herbal patents are first generation migrants or more settled ethnic inventors in the U.S. To tease this out, we employ an exogenous shock to H1B employment visas in the U.S. In 2000, Congress passed the American Competitiveness in the 21st Century Act (AC21) which temporarily increased the quota on H1B visas. In addition to increasing quotas for H1B visas, AC21 also created a visa exemption category that exempted universities and a selected list of other entities from the same quota. To the best of our knowledge, we are the first researchers to use this "exclusion list" to H1B quotas. We exploit these policy changes to estimate a "triple difference" (DDD) model and find that herbal patents are likely to be filed by first generation Chinese/Indian migrants. We also use a novel textual based measure (Google N-grams) to compute the extent of contextual knowledge in each herbal patent and find that Chinese/Indian inventors are more likely to file herbal patents that have a larger proportion of contextual knowledge.

We also present evidence that once codified, knowledge related to herbal patents is initially disseminated through ethnic inventor networks. Additionally, Chinese/Indian inventors are more likely to engage in knowledge arbitrage (i.e. transferring knowledge from their home context to the western labs) and inventors belonging to other ethnic communities engage in knowledge recombination (i.e. combining herbs to other synthetic compounds to create relatively novel formulations).

Our findings contribute to the literatures on inventor mobility and knowledge flows (Rosenkopf and Almeida 2003, Song *et al*. 2003, Oettl and Agrawal 2008, Breschi and Lissoni, 2009), and the role of skilled migrants in knowledge transfer (Kerr, 2008, Foley and Kerr, 2013, Franzoni *et al*., 2014). Our findings also inform the broader literature in strategy and innovation on the microfoundations of knowledge recombination within firms (Fleming 2001, Carnabuci and Operti, 2013, Gruber *et al*., 2013). From the perspective of knowledge recombination, our results points to the possible existence of a complementary relationship between ethnic migrant inventors, (who introduce contextual knowledge to the firm) and non-ethnic inventors (who recombine such knowledge). The rest of the paper proceeds as follows. Section 2 describes two motivating examples and develops theory regarding the production and transfer of contextual knowledge through ethnic links. Section 3 presents the data collection and variable coding process as well as our identification strategy. Section 4 presents results and Section 5 concludes.

Section 2: Motivating Examples and Theory

Two Motivating Examples

In 2015, one of the Nobel Prize winners in in Physiology or Medicine was Tu Youyou from China. She was awarded the prize for "her discoveries concerning a novel therapy against Malaria."[3] Professor Youyou started working on this research around 1967 as part of "Mission 523" or "Project 523", a project launched by Chairman Mao Zedong in response to the loss of life of soldiers to malaria during the Vietnam War. In interviews, Professor Youyou described how she traveled to the southern Chinese island of Hainan to study malaria and then scoured ancient Chinese medicinal text books for remedies, including a book written in 340 BC by Ge Hong, titled A Handbook of Prescriptions for Emergencies.[4] At the onset of her research, the western scientific community had tried around 240,000 compounds to find a cure against malaria, without much success. During her research, Professor Youyou and her team found a brief reference to a novel herb, sweet wormwood, which had been used to treat malaria in China since 400 AD. The research team then extracted an active compound, artemisinin, from wormwood and used the ancient Chinese text books to effectively activate the properties of the compound (e.g. by heating the extract without allowing it to reach boiling point). Since then, artemisinin has been used to cure hundreds of thousands of malaria patients worldwide. Professor Youyou also arranged for the structure of artemisinin to be studied at the Chinese Academy of Sciences in 1975, performed clinical trials in 1977, and published her research in Chinese in the same year. The first article on artemisinin in English was published in 1982. This illustrates the case of contextual knowledge being codified by a local researcher in China and knowledge that did not transfer across borders until much later.

In our second motivating example, we profile Dr. Hari P. Cohli, a researcher at the University of Mississippi in the field of Immunology. He had migrated to Canada to pursue his undergraduate studies at the University of Toronto and subsequently studied and worked at SUNY Buffalo and the Johnson Space Center in Houston. The researchers of this paper interviewed Dr. Cohly on his experiences in filing a United States patent on the medicinal properties of the Indian herb turmeric. At the University of Mississippi, Dr. Cohly came in contact with a plastic surgeon named Dr. S.K. Das, who was about to amputate the leg of a patient, whose wound would not heal because of a condition known as "restenosis," where there is gap between two blood vessels. Dr. Cohly had spent his early years in the Indian city of Agra and had attended Indian herbal medicinal (*Ayurveda*) discourses from Dr. MB Lal Sahab. Dr. Sahab was a religious teacher and a parasitologist who was educated in Edinburgh in parasitology and was the head of Indian Association of Parasitology. He used to conduct these Ayurveda discourses for the *Radhaswami* religious sect in the

[3] http://www.nobelprize.org/nobel_prizes/medicine/laureates/2015/. Website accessed on January 12, 2016

[4] Sources: http://www.bbc.com/news/blogs-china-blog-34451386; http://www.cnn.com/2015/10/06/asia/china-malaria-nobel-prize-tu-youyou/ and http://thewire.in/2015/10/05/how-an-ancient-chinese-text-fought-malaria-and-won-a-nobel-while-india-lags-behind-12381/. Websites accessed on January 12, 2016

Dayalbagh region of Agra, a community of which Dr. Cohly was a member. Dr. Cohly used his contextual knowledge from these discourses and suggested the use of turmeric to heal the wound of the patient. After the patient recovered from his wound and amputation was avoided, Dr. Cohly and Dr. Das conducted a clinical trial at the University of Mississippi and filed a U.S. patent (publication number: US5401504 A), where they claimed *a method of promoting healing of a wound in a patient, which consists essentially of administering a wound-healing agent consisting of an effective amount of turmeric powder to said patient*. This brief motivating example documents the role of an ethnic Indian migrant researcher codifying and transferring the medicinal properties of turmeric to the west.

Theory and Hypotheses

Ethnic migrant inventors and transfer of contextual knowledge across borders

Before we present arguments on why ethnic migrant inventors are likely to play an important role in transferring contextual knowledge across borders, it is important to provide a clear definition of what we mean by contextual knowledge. We draw on Hambrick and Macmillan (1985) who stated that "context refers to the environment and broad organizational milieu in which the innovative attempt is situated" (Hambrick and MacMillan, 1985; page 529). We then draw on the literature in strategy and international business that has documented cross-national variation in context along cultural, linguistic, and other dimensions (Ghemawat 2001, Berry *et al*. 2010) and define contextual knowledge as knowledge that is deeply embedded in its cultural, religious or linguistic context.[5]

We draw on the prior literature on codification of tacit knowledge (Polanyi, 1966, Dasgupta and David 1994, Cowan and Foray, 1997) to argue that ethnic migrant inventors play an important role in transferring contextual knowledge across borders. Arguably, the starting point of this literature is Arrow (1962), who described "invention as the production of information" (Arrow 1962; page 616). Arrow (1962) also described information as a "commodity" and asserted that "the cost of transmitting a given body of information is frequently very low" (Arrow 1962; page 614). Subsequent research has however pointed out several difficulties of knowledge flows across borders, especially if such knowledge is tacit.

Dasgupta and David (1994) define tacit knowledge as the "context which makes focused perception possible, understandable and productive" (Dasgupta and David, 1994; page 493). As an example of tacit knowledge needed for the production of science, the authors talk about "scientific expertise" which is acquired through experience and transferred by demonstration, by personal instruction and by the provision

[5] The literature in global strategy has outlined several dimensions of cross-national variation in context. Variation in cross-national context leads to 'distance' between geographic regions and some of the dimensions of cross-national contextual distance that have been highlighted by prior research include cultural distance (Hofstede 1984), institutional distance (Kostova 1996), and economic distance (Tsang and Yip, 2007). There is also a literature on how cultural and other dimensions of distance lead to poor communication and impedes knowledge transfer (Lin and Germain, 1998; Zhou and Wu, 2010).

of expert services such as advice, consultation, etc. Tacit knowledge could also be thought of as practical knowledge needed to create, use, or adapt new innovations. The concept of tacit knowledge dates back to Polanyi's phenomenology and embodies the idea that many human skills and much of human expertise is dependent on a range of unconscious tacit processes (Polanyi, 1966). Polanyi observed that a skillful performance by an innovator might be achieved by following a set of rules that might not be known to even the person following them.[6] Polanyi also articulated the difficulty of transferring tacit knowledge by prescription, as "no prescription for it exists." Subsequently, scholars writing in the innovation literature such as Dasgupta and David (1994) and Cowan and Foray (1997) have made a strong argument in favor of codifying tacit knowledge, i.e. the process of converting tacit knowledge into messages which can be processed as information. Dasgupta and David (1994) summarize the main benefit of codification of tacit knowledge in that it renders the transmission, verification, storage, and reproduction of information all less costly. They also make an argument for undertaking measures such as granting patents, to ensure strict non-excludability of codified knowledge, i.e. restricting access to those who do not have a right to use it. There is also a well-established literature on how tacit knowledge can be codified using a three step process of model building, language creation, and the writing of messages (Cowan and Foray, 1997).

We draw on the prior literature on tacit and codified knowledge in innovation and argue that ethnic migrant inventors are uniquely positioned to codify contextual knowledge once they migrate to research labs within U.S. firms and universities. If the relevant contextual knowledge is ex ante tacit, as the motivating example on turmeric shows, one could build on the definition of tacit knowledge articulated by Dasgupta and David (1994) and argue that ethnic migrant inventors, prior to the migration, were deeply embedded in the "*context,"* which makes focused perception possible, understandable, and productive. In fact, ethnic migrant inventors were not only embedded in the relevant context, they also have greater access to the relevant "scientific expertise" through closer contact to experts who harbor the contextual knowledge. This makes them ideal candidates to codify contextual knowledge that is ex ante tacit.[7] There is also a possibility that ex ante the relevant contextual knowledge is available in a codified format in the home country of the migrant ethnic inventor. Even in this case, the migrant ethnic inventor is in a unique position to codify this knowledge using the standard language of the codification used in the western research lab (e.g. filing claims within U.S. patent text). This argument relates to the difficulty in translating codified knowledge across contexts. Borjas and Doran (2012) document the poor translation rates of Soviet text

[6] Von Hippel (1994) provides several examples to substantiate this claim and mentions about medical experts who may not be aware of the rules they follow to reach a diagnosis of various systems.

[7] Innovation scholars such as Nightangle (2003) build on Polanyi's phenomenology to argue that neurological hardware of inventors interact dynamically with their physical and cultural environment to generate a range of related but unconscious neural images relevant for tacit knowledge. Given the proximity of ethnic migrant inventors to the "physical and cultural environment" relevant for the contextual knowledge in question, it is likely that they will be in an advantageous position to codify such tacit contextual knowledge when they move to an environment, such as a research laboratory, that supports the codification.

books on Mathematics into English prior to the fall of the Soviet Union. Also, even if the contextual knowledge is codified in the home country of the inventor (e.g. in a native language book), there might be tacit knowledge needed to interpret this codified knowledge, to transfer the knowledge to the west. As an example, to quote Kerr (2008), hindrances to knowledge flows "may result from inadequate access to the informal or practical knowledge that complements the codified details of new innovations" (Kerr 2008; page 518). This leads us to our first hypothesis:

Hypothesis 1: Contextual knowledge is more likely to be codified by ethnic migrant inventors.

Subsequent spread of contextual knowledge through ethnic networks

We next draw on the innovation literature related to the transfer of knowledge through ethnic networks and the literature on transferring sticky knowledge in strategy to argue that contextual knowledge is not only likely to be produced by migrant ethnic inventors, it is also likely that post-codification, contextual knowledge will be initially disseminated through ethnic inventor networks.

The innovation literature (Agrawal *et al*., 2006, Breschi and Lissoni, 2009) has documented that social ties are related to knowledge flows. Agrawal *et al*. (2008) conclude that spatial and social proximity are substitutes in their influence on access to knowledge flows. There is also an emerging literature on the role of ethnic inventors and Diaspora in facilitating knowledge transfer. Kerr (2008) notes that ethnic scientific networks are important for short-term technology transfer from the U.S. In her study of Chinese and Indian engineers and entrepreneurs in Silicon Valley, Saxenian (1999) outlines the role of trust and reciprocity in transferring knowledge among members of the ethnic community. In a study of the Indian software industry, Nanda and Khanna (2010) find that Diaspora networks may serve as substitutes for local institutions in helping entrepreneurs outside the software hubs access knowledge. A related paper is by Foley and Kerr (2013), who study the impact ethnic inventors have on the global activities of U.S. firms. The authors find that a 10 percentage point increase in the share of innovation by individuals of a particular ethnicity is associated with a 1 percentage point increase in the share of multinational affiliate activity in countries related to that ethnicity. Agrawal *et al*. (2011) find that inventors based in India who work for multinational firms disproportionately cite the Indian Diaspora than do those who are employed by the same firm but are based at facilities in other countries. Almeida *et al*. (2014) find evidence of intra-ethnic citations in the U.S. semiconductor industry. Docquier and Rapoport (2012) provide a useful overview of this literature.

In this paper, we argue that while social and ethnic proximity could determine knowledge flows for a broad set of knowledge, it is particularly relevant for contextual knowledge. To make this argument, we draw on the literature on impediments to transferring sticky knowledge (Von Hippel 1994, Szulanski 1996, Jensen and Szulanski 2004). Szulanski (1996) outlined several impediments to the transfer of knowledge including causal ambiguity and the recipient's lack of motivation. As Szulanski (1996) states, causal ambiguity can result from imperfectly understood idiosyncratic features of the context in which the knowledge is put to use. In the case of contextual knowledge transferred by an ethnic inventor, it is likely that inventors from other ethnicities might suffer from causal ambiguity in further working with such knowledge. That might lead to low motivation on part of inventors from other ethnic communities to work on such knowledge, at least initially. This leads to our second hypothesis:

Hypothesis 2: Once codified, c*ontextual knowledge is initially more likely to spread through ethnic inventor networks.*

Recombination of contextual knowledge — microfoundations

Next we theorize about how contextual knowledge gets recombined after its transfer to the host western firm and draw on the strategy and innovation literature related to the microfoundations of knowledge recombination.

There is a rich tradition of studying knowledge recombination across 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 individuals play within the firm with respect to (w.r.t.) knowledge recombination. This tradition dates back to Allen (1977) and is framed by Fleming (2001) as the process of recombinant search of individual inventors.

In fact, Fleming (2001) states, "inventors constantly import previously untried components from outside the extant made world, for example *the use of medicinal substances from tropical jungles"* (italics added by current authors) (Fleming 2001, page 119). Building on March (1991) and Cohen and Levinthal (1990), Fleming (2001) frames recombinant search as either "distant" (when the inventor tries completely new components or combinations) or as "local recombinant search" (when the inventor recombines from a familiar set of technology components). In the subsequent literature, Carnabuci and Operti (2013) have described these two recombination strategies as "recombinant creation" (i.e. creating recombinations new to the firm) and "recombinant reuse" (i.e. reconfiguring combinations already known to the firm) respectively.

We build on this literature to theorize that in the case of contextual knowledge, ethnic migrant inventors are likely to engage in knowledge arbitrage, while inventors of other ethnic backgrounds are likely to engage in recombinant creation. In our setting, we frame knowledge arbitrage as the ethnic migrant inventor appropriating knowledge from a prior context (i.e., the ethnic migrant home country context) and codifying that knowledge in a new context (i.e. the western firm).

Building on the prior literature on skilled migration, we argue that the ethnic migrant inventor is well suited to arbitrage her unique contextual knowledge once she moves to the western research entity. In fact Franzoni *et al*. (2014) state the following — "because knowledge is largely tacit and embedded in individuals, migrant scientists can arguably be exceptionally productive because mobility places them in a position of arbitrage" (Franzoni *et al*., 2014; page 2).

We also theorize that inventors of other ethnic backgrounds are better suited to engage in knowledge recombination, especially recombinant creation. This relates to the construct of knowledge diversity of individual inventors. In the strategy literature, Cohen and Levinthal (1990) state that knowledge diversity facilitates the innovative process of individual inventors, by helping them make novel associations and linkages w.r.t to the problem they are attempting to solve. In a similar vein, Ahuja and Lampert (2001) have shown that knowledge diversity helps individuals engage in a radically different approach to solving a technological problem. Extending this argument, Carnabuci and Operti (2013) theorize that knowledge diversity helps individual inventors engage in recombinant creation. In the case of contextual knowledge being transferred to a western firm, inventors of non-ethnic backgrounds (e.g. inventors of western ethnic backgrounds) are likely to have greater knowledge diversity compared to ethnic migrant inventors. While ethnic migrant inventors are plausibly well versed in contextual knowledge, inventors of other ethnicities are plausibly better versed in knowledge codified in the west. If such inventors of other ethnicities are additionally exposed to contextual knowledge, they could engage in recombinant creation. In other words, inventors of other ethnic backgrounds are likely to have greater breadth of organizational knowledge search (Paruchuri and Awate, 2016 compared to ethnic migrant inventors. This leads to our third hypothesis:

Hypothesis 3: While ethnic migrant inventors are likely to engage in knowledge arbitrage, inventors of non-ethnic backgrounds are likely to engage in recombination of contextual knowledge.

Data, Variables, and Identification Strategy

To study the relationship between contextual knowledge and ethnic inventors, we use a unique dataset of herbal patents filed in the U.S. Herbal patents are an appropriate empirical setting to study the production and transfer of contextual knowledge for several reasons. China and India together compose around a fifth of the world's known plant species.[8] Furthermore, for centuries, the two countries have accumulated extensive knowledge on these plant species as part of distinct medical systems (*Ayurveda*, *Unani*, *Siddha*, *Yoga*, and TCM, or Traditional Chinese Medicine). Additionally, there is a large population of Chinese and Indian migrant knowledge workers in the west, and they are the largest beneficiaries of temporary work

[8] Source: The world resources 2005. (http://www.undp.org/content/dam/undp/library/Environment%20and%20Energy/biodiversity/wrr05_lores.pdf)

visas to the U.S. This represents an opportunity to test whether or not contextual knowledge of herbal medicine is patented in the west by migrant Chinese and Indian inventors and whether or not initial citations come from ethnic inventors.

Unique dataset of herbal patents

Starting with the entire universe of USPTO patents, we searched for and identified 2,060 herbal patents filed between 1977 and 2010 using Thompson Innovation and LexisNexis TotalPatents. We categorized patents as herbal if they contain at least one herb name and its use. Our search process consisted of three iterative steps. First, we performed keyword searches for patents that contain herb names in either the abstract or title. Second, we searched for herb related patents within relevant patent classification categories. Third, we collected patents from the Traditional Chinese Medicine (TCM) database with priority in the United States. The process was completed by manually reading through and categorizing patents as herbal patents or not. We detail the search process below.

First, we obtained a list of 52 herbs, their common names, and their scientific names from the National Center for Complementary and Alternative Medicine (NCCAM) website. Using these herbs, we searched Thompson Innovation and LexisNexis TotalPatents for USPTO patents containing any of these herbs in the abstract or title. In addition to the 52 herb names included in the NCCAM website, we searched for additional herb names within the identified herbal patents. We extracted herb names from each herbal patent and concatenated these to form a list of 499 herb names. The most frequent herbs were "soybean" and "Soy," which collectively account for about 5 percent of the patent-herb pairs. The total number of patent-herb pairs is greater than the total number of herbal patents, because one patent can contain multiple herbs. Table 1 shows the 10 most frequent herbs in our database.

[Insert Table 1 Here]

Next we performed a classification search. We used both International Patent Classification (IPC) and the US Patent Classification (USPC) schemes, and in particular, IPC class A61K36+, and USPC classifications 424/725 and 514/783. This IPC class was introduced in 2002 by a Committee of Experts at IPC Union for linking Traditional Knowledge Research Classification (TKRC) with IPC as a part of the work by the World Intellectual Property Organization Traditional Knowledge (WIPO-TK) Task Force. Finally, we used the Traditional Chinese Medicine (TCM) database to augment our dataset, and read patent abstracts to further validate our list. From the TCM database, we collected all patents with U.S. priority and appended this to our existing dataset. We manually read through the titles and abstracts of our patents to identify other herb names and their usages. The resulting list of patents consists of 2,060 patents filed between 1977 and 2010.

Generating control patents

To create a control patent dataset, we used a matching technique similar to Jaffe *et al*. (1993) but in the base case, we went one step further by controlling for textual similarity of patents. As Thompson and Fox-Keane (2005) pointed out, patent classification codes may be too broad to serve as adequate controls. In our case, there is an additional challenge since many herbal patents belong within the same patent class. Ideally, we would like to control for the specific medicinal usage of the patent such as curing cancer and to do this, we therefore created a control group based on the textual similarity of patent usage. For each of our herbal patents, we document the medical application of the patent. Using Google Patents, we searched for patents with the same medical application. For instance, a typical search term would consist of phrases such as "treat gastrointestinal inflammation." Additionally, we controlled for the application date.

Identification strategy and variables

We proposed three hypotheses: (1) that contextual knowledge is more likely to be codified by ethnic (Chinese/Indian) migrants; (2) that once codified, contextual knowledge is initially more likely to spread through Chinese/Indian inventor networks and (3) while ethnic migrant inventors are likely to engage in knowledge arbitrage, non-ethnic inventors are likely to engage in knowledge recombination. To test the first hypothesis, we run a triple differences model with an indicator for Chinese/Indian migrants on a patent as the dependent variable, and see if this variable changes as the flow of Chinese/Indian migrants to the U.S. changes. To test the second hypothesis, we use the fraction of forward citations by Chinese/Indian inventors to herbal patents as the dependent variable, and see whether contextual knowledge spreads disproportionately through ethnic ties, compared to matched control patents. To test the third hypothesis, we code whether or not herbal patents are 'recombined' i.e., whether or not they include synthetic compounds in addition to herbs and then test whether recombined patents are filed by non-ethnic inventors. In the following section, we describe our natural experiment, our variable definitions, and the empirical specifications.

Natural experiment: The H1B visa shock and excluded entities

Even if herbal patents are more likely to be filed by ethnic Chinese and Indian inventors compared to matched control patents, it is unclear whether these inventors are first generation migrants or more settled ethnic inventors in the U.S. If being embedded in the context is indeed related to codification of contextual knowledge, we would expect to see first generation migrants disproportionately writing herbal patents. Therefore, we aim to test whether it is the stock or flow of ethnic Chinese and Indian inventors to the U.S. that is driving herbal patent filing.

Towards this goal, we utilize an exogenous shock to Chinese/Indian immigrants in the U.S.[9] In 1998 and 2000, Congress introduced two laws that significantly increased the flow of skilled immigrants. As a result of these two legislations, the number of H1B visas increased from 65,000 in 1998 to 115,000 in 1999, up to 195,000 again in 2001, and back down to 65,000 in 2004.[10] The laws were introduced in response to the increased demand for IT professionals during the dot com bubble. Therefore, the flow of new migrants is plausibly exogenous to filing herbal patents, as most of the workers are hired in IT-related occupations. We focus on Chinese and Indian inventors because they are the two largest groups to receive H1B visas: workers from India comprise the majority of H1B recipients, followed by workers from China. Figure 1 outlines the cap of H1B visa issuances over time. In summary the H1B visa quotas were elevated between 1999 and 2003. In the base case, we consider 2000–2004 as the treatment period (*TREAT*) given that migrants moving to the U.S. would probably need at least a year before they could start filing patents. In robustness checks we relax this constraint.

[Insert Figure 1 Here]

Certain firms were exempted from the visa cap under the same regulation. Workers who work "(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"[11] could hire as many employees as they wanted to through the H1B visa. This allows us to compare the differential effect of the visa cap increase by comparing the cap-subject and cap-exempt groups of patenting entities.

We use a triple-differences model to estimate the causal impact of Chinese/Indian migrant inventor flows on the patenting of contextual knowledge. In effect, we are comparing two differences-in-differences estimates. Our first difference-in-difference (DD) comes from comparing patent authorship for cap-exempt and cap-subject groups during the shock period. We repeat this DD for herbal and control patents, and compare the coefficients to obtain the triple difference (DDD). The last difference step controls for any non-parallel trends that might be present across cap-subject vs. cap-exempt firms.

We are not looking at whether an increase in an inventor group increased the number of herbal patents. Rather, we are showing that herbal patents have more Chinese/Indian inventors than matched control patents, and that this gap changes in response to changes in immigration patterns. Given that the immigration quotas went up during the period of treatment, if the gap increases, it would be consistent with our hypothesis that the creation of contextual knowledge is related to an increased flow of first generation ethnic immigrants.

[9] A similar shock has been used by Kerr and Lincoln (2010).

[10] The American Competitiveness and Workforce Improvement Act (ACWIA) was passed in 1998, and the American Competitiveness in the 21st Century Act (AC21) was passed in 2000. AC21 has a clause that also retroactively increased the quota for 1999 and 2000, past the 115,000 cap set by the ACWIA.

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

Dependent variables

Inventor ethnicity

Patent documents do not record ethnicities of inventors, but we were able to predict the most likely ethnicities based on linguistic cues left by their names. 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.[12] 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. The appendix reports correlations across our measures and other established measures of ethnicity classification (Appendix Tables A1 A2). We find correlations mostly above 0.9 for all of our ethnicity measures. Once the program receives a name as an input, it returns the statistically most likely ethnicity of that name based on a standard machine-learning algorithm[13].

As our main dependent variable for testing Hypothesis 1, we use an indicator denoting whether the patent contains any ethnic (i.e. Chinese or Indian) inventors (*ETHNIC INVENTOR*). If a patent contains any Chinese or Indian inventors, the variable is coded 1, and 0 otherwise. For robustness checks we use a second measure of ethnic inventors' patenting activity, the fraction of ethnic inventors on a patent. For each patent, we sum the number of inventors categorized as either Chinese or Indian, and we divide this number by the total number of inventors on that patent. While the indicator variable captures the probability of finding any ethnic inventors in a patent, the ethnic fraction measure captures the average ethnic inventor activity for any given patent. Furthermore, we also categorized inventors as "European" if their first and last names are classified as any of the following 14 ethnicities: French, Czech, Italian, German, Jewish, Swedish, Ukrainian, Spanish, Portuguese, Swiss, Danish, Irish, Greek, or Russian.

Patent citations

To create the dependent variable for Hypothesis 2, we collect all patents citing our herbal and control patents, and code the ethnicities of the inventors in each citing patent. For each patent, we calculate the fraction of citing patents that list a Chinese/Indian inventor (*FRACTION OF CITATIONS ETHNIC*). Thus, for each patent, we can calculate how many citations in a given month are by Chinese or Indian inventors. Here, patent citations are a proxy for knowledge flows. Given that a fraction of our citing patents are added

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

[13] In particular, the program uses the Maximum Entropy classifier. This algorithm estimates parameters on linguistic features that maximize the posterior likelihood of a name being classified into ethnic categories.

by patent examiners (Alcacer and Gittelman 2006), we discard these in our analysis. In the base case, we also discard any citations that occur within the same firm, but results are robust to adding these citations back.

Recombined (knowledge)

To create the dependent variable for Hypothesis 3, we code herbal patents as *RECOMBINED* if the patent text contains reference to synthetic non-herbal formulations in addition to referencing herbs. The dependent variable (*RECOMBINED)* is coded equal to 1 if the herbal patent contains one or more synthetic compound in addition to containing one or more herbs.

Secondary variables

Measuring "contextual knowledge" using Google N-grams

For each herb name in our dataset, we obtained the Google Ngram scores[14] between 1977 and 2008. The Google Ngram viewer allows the user to see how often a keyword appears across time in books digitized by Google. Specifically, for a given keyword, it returns the number of times that keyword appeared over the total number of words in a given year. Furthermore, Google allows us to customize which corpus of books are to be used for the search. We searched the default American English corpus which consists of books predominantly in the English language that were published in the United States. The variable $CONTEXT$ measures the negative log frequency of the herb name in the American English language corpus. This variable captures, in a sense, how "out of context" the herbs are from the standpoint of inventors in the U.S. Low frequency words would thus be further away from the U.S. context. The rankings across years are more or less consistent, and we use the Ngram word frequency counts for the years 1977 (year when our sample starts) and 2008 (year when Google Ngram data ends). In the base case, we used the frequency counts for the year 2008. Inspection of the Ngram data confirm our belief that contextual information can be captured by frequency counts. Herbal ingredients in forms such as apples, tobacco, pine, and sage are used frequently in the English language, and thus get low scores on the variable *CONTEXT*. On the other hand, herbs such as *Aeginetia* (forest ghost flower, native to India) or *Fructus tribuli* (Chinese herb *Bai Ji Li*) appear less frequently, and get high scores on the variable *CONTEXT*.

Classifying assignees

We categorized assignees into three broad groups: Individuals, Universities/Affiliated Research Organizations/Nonprofits, and Others. We matched each assignee name to an identifier using CapitalIQ to

[14] Source: https://books.google.com/ngrams

clean typographical errors. We then manually sorted through the list of patents to categorize patents into one of the three groups of assignees.

The list of Universities/Affiliated Research Organizations/Nonprofits was further tailored to our specific case. We obtained a list of H1B visa cap-exempt employers from a 3rd party online employment entity.[15] The online list contains 12,479 employers who have been categorized as exempt from the H1B visa cap. We matched these employers to our list of assignees, and further searched for "university" and "college" to construct a list of assignees that we expect to be exempt from the H1B visa cap (*CAP*). Out of the 998 total number of unique assignees in our herbal patent sample, 63 unique assignees are exempt from the H1B visa cap. Out of the 1433 total number of unique assignees in our matched control patent sample, 57 unique assignees are exempt from the H1B visa cap. Table 2 lists a partial list of cap-exempt assignees.

[Insert Table 2 Here]

Empirical specifications

To test Hypothesis 1, i.e. contextual knowledge is more likely to be codified by Chinese/Indian migrant inventors, we measure whether an increase in the flow of Chinese/Indian inventors differentially affects the fraction of Chinese/Indian inventors writing herbal and control patents. Our identification comes from comparing the difference in difference results from the H1B visa shock for our two groups of entities: cap-subject and cap-exempt. Towards this, we estimate the following regression equation:

$$\boldsymbol{ETHNIC\ INVENTOR = \alpha + \beta_1 HERB + \beta_2 CAP + \beta_3 TREAT + \gamma_1 HERB \times CAP + \gamma_2 CAP \times TREAT + \gamma_3 HERB \times TREAT + \delta HERB \times CAP \times TREAT + \xi X + \varepsilon} \quad \textbf{(1)}$$

Our dependent variable (*ETHNIC INVENTOR*) is an indicator for whether a patent has an ethnic (i.e. Chinese/Indian) inventor. The variables $HERB, CAP, TREAT$ are dummies for whether the patent is an herbal patent, whether the assignee is subject to the H1B cap, and whether the patent was filed between the treatment period of 2000 and 2004. We include various controls, denoted X, for patents (e.g., time trends, citation count to control for underlying patent quality, inventor count, etc.).

Here, the β coefficients capture the time-invariant difference between herbal and control patents (β_1), time invariant differences between cap-subject and cap-exempt patents(β_2), and the changes in inventors over time (β_3). The two-way interaction terms γ capture the time invariant characteristics of the herbal patents by cap-subject assignees (γ_1), changes in cap-subject patents over time (γ_2), and changes in herbal patents over time (γ_3). Finally, the triple interaction term δ captures whether or not increase in

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

immigration flows of Chinese/Indian inventors to cap-subject firms, leads to observable differences in herbal patents being filed by ethnic inventors.

To test Hypothesis 2, that is once codified, contextual knowledge is initially more likely to spread through Chinese/Indian inventor networks, we run a regression with the following specification

$$\boldsymbol{FRACTION\ OF\ CITATIONS\ ETHNIC = \alpha + \beta_1 ETHN + \beta_2 HERB + \beta_3 TIME + \gamma_1 ETHN \times HERB + \gamma_2 TIME \times HERB + \gamma_3 TIME \times ETHN + \delta TIME \times HERB \times ETHN + \xi X + \varepsilon} \quad \textbf{(2)}$$

Here $HERB, ETHN,$ denote dummy variables for herbal patents, and whether or not the herbal/control patent has an ethnic inventor, and $TIME$ denotes the time in months it took for the citation event to occur. Our dependent variable (*FRACTION OF CITATIONS ETHNIC*) is the fraction of forward citations in a given month that have any Chinese/Indian inventors.

Our goal is to see whether citations to herbal patents disproportionately spread through ethnic networks, even when controlling for factors such as ethnicity of the inventors filing the patent and medical application. The β coefficients denote time invariant differences between patents by Chinese/Indians and patents without (β_1), time invariant differences between herbal patents and control patents (β_2), and changes in citations over time (β_3). The interaction terms γ capture the time invariant characteristics of herbal patents by Chinese/Indian inventors (γ_1), changes in citations for herbal patents (γ_2), and changes in citations for patents by Chinese/Indian inventors (γ_3). Finally, the coefficient of interest δ shows how other Chinese/Indian inventors cite herbal patents by Chinese/Indian inventors over time.

To test Hypothesis 3, which states that while ethnic migrant inventors are likely to engage in knowledge arbitrage, inventors of other ethnic backgrounds are likely to engage in recombination of contextual knowledge, we have to measure whether patents are the result of knowledge recombination or arbitrage. To do this, we categorize each herbal patent into whether it is a *pure* herbal application or a recombined herbal application. We code herbal patents as *RECOMBINED*, if the patent text contains references to synthetic, non-herbal formulations in addition to referencing herbs. If first generation immigrants are arbitraging their contextual information once they migrate to the U.S., we should see an increase in the probability of an herbal patent being a pure herbal patent (i.e. comprising only herbs and no synthetic compounds) during the period of the shock. Using the *RECOMBINED* variable, we test whether there is a significant association between the number of Chinese/Indian inventors on an herbal patent and what type of knowledge is created (i.e. recombination or arbitrage). Furthermore, we test what effect the H1B visa shock had on pure herbal patents versus herbal patents with both herbs and synthetic formulations. We run the following regression equation:

$$\boldsymbol{RECOMBINED = \alpha + \beta_1 TREAT + \beta_2 CAP + \gamma TREAT \times CAP + \varepsilon} \quad \textbf{(3)}$$

The dependent variable (*RECOMBINED)* is coded equal to 1 if the herbal patents contains synthetic compounds in addition to containing one or more herbs. $TREAT$ is an indicator denoting whether the observation was made during the treatment period, and CAP is an indicator denoting whether the assignee of the observation was subject to the visa cap. The β coefficients denote the change in synthetic patenting over time (β_1), and the time invariant difference in synthetic patenting between cap-subject and cap-exempt patents (β_2). We also compare whether the treatment effect γ is different for patents with and without Chinese/Indian inventors, through sub-sample analyses. We should also observe an increase in the overall rate of recombination during the visa shock period due to the increase in knowledge diversity (β_1). Hypothesis 3 predicts that since ethnic migrant inventors engage in arbitrage, we should see a decrease in recombination for companies that employ larger number of Chinese/Indian migrant inventors during the visa shock period (γ).

Results

Summary statistics for herbal and control patents

In our results, we expect to see more Chinese/Indian inventors on herbal patents, even after controlling for selection into medicinal application area. Our control group patents are selected so that they have the same medicinal application as herbal patents, but without herbs. If we see that the participation of Chinese/Indian inventors on herbal patents is significantly different from our control group, we can argue that herbal patents are more likely to be codified by ethnic inventors with contextual knowledge, conditional on the medicinal application. Our sample (i.e. herbal and control patents together) contains a total of 9,068 unique inventors, 923 of which have Chinese names, and 566 of which have Indian names. Table 3 presents summary statistics for our control and herbal patent groups, broken down by cap-exempt and cap-subject sub-groups for both control and herbal patents. In addition, Table 4 presents summary statistics and t-tests for control and herbal patents, *on average*, across both cap-exempt and cap-subject sub-samples.

[Insert Table 3 Here]

Table 4 indicates that herbal patents overwhelmingly have more Chinese/Indian inventors. There are more herbal patents that have Chinese/Indian inventors than control patents and this difference is statistically significant. The average number of Chinese/Indian inventors on herbal patents is 0.676 per patent, which is more than two times that of control patents, which have 0.333 per patent and this difference is statistically significant. There is no statistically significant difference in the average number of European inventors, with 1.946 per control patent, and 1.875 per herbal patent. We see similar patterns for the fraction of inventors that are Chinese/Indian as well.

[Insert Table 4 Here]

We observe some overlap in the assignees of herbal and control patents. There are a total of 2,729 assignees in our dataset, where herbal patents have a total of 1,215 unique assignees, and control patents have 1,703 unique assignees. Overall, there is significant overlap in the largest assignees of both herbal and control patents, indicating assignees that have many herbal patents also have many non-herbal patents. There are 189 assignees that have written both herbal and control patents, and these assignees write 33 percent and 19 percent of herbal and control patents respectively, over half the patents in our dataset. Firms such as SmithKline Beecham (GSK) and Sunovion Pharmaceuticals have only non-herbal patents in our dataset, while firms such as Coty and Johnson & Johnson only have herbal patents. Table A3 lists assignees with the most patents in our dataset.

Non-herbal patents have a larger number of citations compared to herbal patents and this difference is statistically significant. However, the number of European inventors and the fraction of cap-subject and cap-exempt assignees matches well for our control and herbal patent groups, despite the fact that we did not explicitly control for assignees. Therefore, we believe the matching successfully controls for the medicinal application area usage of the patents.

Testing hypothesis 1 — triple differences estimation

Hypothesis 1 stated that contextual knowledge is more likely to be codified by ethnic migrant inventors. To test this, we present results from estimating equation (1) in Table 5 using an indicator for whether the patent has ethnic Chinese/Indian inventors as the main dependent variable. Across different models, we control for time-specific effects using a time trend, number of forward citations, and number of inventors. Standard errors are clustered at the assignee (employer) level. The triple interaction term δ captures the effect of the increase in the immigration flow of Chinese/Indian inventors on the filing of herbal patents.

[Insert Table 5 Here]

Table 5 indicates that there is a significant, time-invariant difference between herbal and control patents in terms of their likelihood of having an ethnic Chinese/Indian inventor (β_1), echoing our t-test results. Furthermore, there is an increase in Chinese/Indian inventors filing herbal patents during the visa shock treatment (β_3). Table 5 also documents that the treatment effect (δ) is positive and significant, indicating that an increase in the flow of Chinese/Indian migrants to the U.S. is related to a greater likelihood of Chinese/Indian inventor names on herbal patents compared to control patents, in cap-subject firms. The baseline fraction of patents with Chinese/Indian inventors is 0.216, and the effect of the treatment is 0.236 in the baseline model (column 1). This indicates that there is more than a twofold increase in the likelihood of observing ethnic Chinese/Indian inventor names on herbal patents for the cap-subject sample, after the treatment. The effects are significant controlling for time fixed effects, citation counts, and total number of inventors. Furthermore, the results are robust to using the fraction of Chinese/Indian inventors as the dependent variable. Robustness checks reveal there is no statistically significant effect for using count or fraction of European inventors on a patent. Note that since the majority of our patents have unique assignees, the use of assignee fixed effects will severely limit the available variation (only a fifth of our herbal-control patent pairs have assignees with more than one patent) and may introduce sample selection issues.

Graphically, we can plot how the cap-exemption affected herbal patenting participation by Chinese/Indian inventors over time. For each year, we create a subset of our data (herbal and control patents) using only patents from that year and run the following regression.

$$\boldsymbol{ETHNIC\ INVENTOR = \alpha + \beta_1 HERB + \beta_2 CAP + \gamma_1 HERB \times CAP + \xi X + \varepsilon} \qquad \textbf{(4)}$$

Figure 2 plots the effect of the visa cap on the likelihood of observing Chinese/Indian inventors in herbal patents (γ_1) over time. In other words, we plot the mean difference in the likelihood of observing Chinese/Indian inventors in herbal patents across the cap-subject and cap-exempt sub-samples, over time. The effect sizes are normalized so that zero is the average of the pre-treatment treatment effects. We see that during the treatment period, the difference in the likelihood of observing Chinese/Indian inventors in herbal patents across the cap-subject and cap-exempt sub-samples increased to a statistically significant level.

[Insert Figure 2 Here]

Additional tests for hypothesis 1

We also ask whether Chinese and Indian migrants do indeed create patents that contain more contextual knowledge. The main obstacle to tackling such a question is measurement. There is no direct measure for contextual knowledge embedded in a patent and we turn to linguistic traces of the context in the herb names to construct a measure of contextual knowledge for each herbal patent.

For each patent, we collect the names of all herbs mentioned in the patent. Then, for each herb collected, we calculate its empirical frequency in the default American English language corpus. We measure the extent of contextual knowledge in an herb using the negative log frequency of the herb name in the American English language corpus. We expect to see more Chinese/Indian inventor names in patents about herbs uncommon in the American English language corpus. Towards this we run the following regression.

$$\boldsymbol{ETHNIC\ INVENTOR = \alpha + \beta_1 CONTEXT + \xi X + \varepsilon} \qquad \textbf{(5)}$$

Since each patent contains multiple herbs, in the base case, we construct the variable *CONTEXT* using the average of the values for negative log frequency for each herb contained within a patent. As a robustness check, we construct the variable *CONTEXT* using the negative log frequency of the most contextual herb (i.e., the most infrequent herb). Averaging, i.e. the first method, allows us to capture the relationship between inventor ethnicity and the extent of contextual knowledge of all herbs in the patent. Considering the most infrequent herb allows us to capture the association between inventor ethnicity and the rarest herb in the patent. In the base case, we report results using the average context of herbs on a patent, and the results using the most contextual herb are reported in appendix tables A4–A5. Here, β_1 measures the association between contextual knowledge and the likelihood of observing ethnic inventors on a patent. An increase in the contextual knowledge is associated with a $\beta_1$increase in the probability of having a Chinese/Indian inventor on a patent.

Table 6 presents estimation results for equation (5). We see a positive and significant association between the extent of contextual knowledge contained within an herbal patent and the likelihood of observing Chinese/Indian inventors on the patent. Compared to the baseline likelihood of observing Chinese/Indian inventors on a patent, a one standard deviation increase in contextual knowledge increases the likelihood of observing an ethnic Chinese/Indian inventor by 26 percent (details of this computation provided in footnotes of Table 6). For example, patents about *Artemesia* (related to the Chinese wormwood herb profiled earlier) are 26 percent more likely than patents using St. John's wort (a cosmopolitan invasive weed, that has spread to temperate regions across India, China, Canada, Africa, and the United States)[16] to have Indian or Chinese inventors. This effect is robust to controlling for time fixed effects, citation counts, and inventor counts.

[Insert Table 6 Here]

Furthermore, our setting allows us to look at how an inflow of migrant Chinese/Indian inventors affects this relationship between the extent of contextual knowledge of herbal patents and the likelihood of observing Chinese/Indian inventors on the patent. We also run the following specification

$$\boldsymbol{ETHNIC\ INVENTOR = \alpha + \beta_1 CONTEXT + \beta_2 TREAT + \gamma CONTEXT \times TREAT + \xi X + \varepsilon} \quad \textbf{(6)}$$

Here $CONTEXT$ is defined as above, and $TREAT$ denotes the visa shock period. β_1 measures the time invariant association between context and ethnic inventor authorship. β_2 measures the time effect of the visa shock on ethnic inventor authorship. Finally, γ measures the differential association between the likelihood of observing Chinese/Indian inventors on the patent and patents containing infrequent herbs.

[16] Source: https://en.wikipedia.org/wiki/Hypericum_perforatum. Website accessed on December 9 2016.

Table 7 presents estimation results for equation (6). Column 1 of Table 7 shows that a one standard deviation increase in contextual knowledge contained within an herbal patent is associated with a 20 percent and 35 percent increase in the probability of observing Chinese/Indian inventors in the non-treatment and treatment time periods, respectively (see footnotes of Table 7 for detailed calculations). There is a positive time effect of the visa shock (β_2) as expected. The coefficient on the interaction term (γ) is almost as large as β_1, indicating the association between the extent of contextual knowledge in the herbal patent, and the likelihood of Chinese/Indian inventor names on the patent, grew stronger during the treatment period. This is consistent with our hypothesis that contextual knowledge is more likely to be codified by ethnic migrant inventors.

[Insert Table 7 Here]

Testing hypothesis 2

Hypothesis 2 posited that once codified, contextual knowledge is initially more likely to spread through ethnic inventor networks. We observe how the ethnicities of the inventors that cite herbal patents and matched control patents in our sample change over time. For each of the herbal and control patents in our sample, we obtain information about all patents that cite the focal patent including publication dates and inventor ethnicities. Figure 3 plots how the fraction of forward citations that have any Chinese/Indian inventors changes over time. Each point on the diagram corresponds to the fraction of citations by patents with any Chinese/Indian inventors, for one of our focal patents, at a given month since publication. Each point is grouped by whether the focal patent does/does not have Chinese/Indian names, and is/is not an herbal patent. We see a strong propensity to cite within ethnic groups: Chinese/Indian inventors are more likely to cite patents filed by other Chinese/Indian inventors. The question then is whether this effect is stronger in the herbal patent sub-sample compared to the control sub-sample. We test this hypothesis below.

[Insert Figure 3 Here]

Identification of this effect comes from comparisons with the matched control group we previously created. We compare whether herbal patents are cited disproportionately more by Chinese/Indian inventors compared to matched control patents, and whether this effect persists over time. Note that we would expect Chinese/Indian inventors to cite other patents filed by Chinese/Indian inventors for various reasons, notably because of social network ties. We can see whether there is a disproportionately larger ethnic ties effect for herbal patents by differencing it with the ethnic ties effect in the control group.

Table 8 presents the results from estimating equation (2) using OLS. Standard errors are clustered at the original patent level. The baseline model in column 1 suggests that having any Chinese/Indian inventors on a patent is associated with a 56 percent and 82 percent greater chance of being cited by other Chinese/Indian inventors for control and herbal patents, respectively, compared to a control patent filed by non-Chinese/Indian inventors filed within the first month of publication of the focal patent. Chinese/Indian citations are slightly increasing over time. Most importantly, an herbal patent with a Chinese/Indian inventor has a 17 percent higher probability of being cited by other Chinese/Indian inventors than similar control patents. One year after publication, the probability of Chinese/Indian inventors citing an herbal patent by other Chinese/Indian inventors decreases by about 2 percentage points, indicating that non-Chinese/Indian inventors are more likely to cite such patents over time.

[Insert Table 8 Here]

Testing hypothesis 3

Hypothesis 3 stated that while ethnic migrant inventors are likely to engage in knowledge arbitrage, inventors of other ethnic backgrounds are likely to engage in recombination of contextual knowledge and is tested using specification (3). Table 9 presents our results using OLS and clustered standard errors, clustered at the level of assignee.

[Insert Table 9 Here]

Column 1 and 2 of Table 9 use the full sample of herbal patents, while columns 3 and 4 use patents with and without any Chinese/Indian inventors, respectively. The positive and significant coefficient on the visa shock dummy (TREAT) indicates that the treatment period coincides with an increase in overall recombination. This effect is positive and significant for all specifications. Given that the point estimate of the interaction term γ is negative and statistically significant across models 1–3, we conclude that the H1B visa shock *decreased* the use of synthetic compounds within herbal patents, for the cap-subject sub-sample. Furthermore, as the point estimate of γ in column 3 suggests, the effect seems to be driven by patents with Chinese/Indian inventors. This suggests in the cap-subject sub-sample, during the treatment period, more pure herbal patents (i.e. patents with only herbs and no synthetic compounds) were being filed. We interpret this evidence as suggestive of the fact that first generation migrants were arbitraging contextual knowledge. Additionally, it suggests that inventors with non-Chinese/Indian names are relatively more likely to engage in recombination, i.e. filing recombined patents.

Robustness checks

Placebo test. Serial correlation in the outcome variable may bias standard errors in difference in differences estimates causing us to underestimate the standard errors (Bertrand et. al., 2004). We follow Chetty *et al*. (2009) and run a permutation test to study whether our estimates suffer from such biases. Intuitively, the permutation test calculates the probability that we will see a similar effect size when the treatment groups and treatment periods are randomly selected. Here, the null hypothesis of the test is a null treatment effect ($\delta = 0$.)

From our sample of herbal and control patents, we randomly select a group of 2,060 patents to be our placebo herbal patents (treatment group), and also randomly select a consecutive 5-year period to be our placebo H1B visa shock period, and we run specification (1) as above, saving the coefficient on the triple differences (DDD) estimate each time. We also randomize the assignment of patents with cap-subject firms. We repeat this process for 11,600 random placebo *triplets*. We select the random placebo triplets based on three dimensions — assignment of treatment (done 20 times each), assignment of time period (done 29 times each for the 29 different possible 5-year time periods), and assignment of cap-subject/cap-exempt status (done 20 times each) for a total of $20 \times 20 \times 29 = 11{,}600$ random placebo triplets. We plot the cumulative distribution function of the DDD coefficients (δ). Our permutations do not suffer from serial correlation in outcomes due to random assignment. Similar to a p-value, if the visa shock positively affected herbal patenting behavior, we would expect our coefficient to be larger than random, and thus appear near the upper right tail of the cumulative distribution function. Results are reported in Figure 4.

[Insert Figure 4 Here]

As Figure 4 indicates, the point estimate for δ that we observed in the fully specified model (column 4) of Table 5 (i.e. 0.213) is likely to be observed less than 10 percent of the time by chance, boosting our confidence in the results. Similarly, when we use the fraction of ethnic Chinese/Indian inventors as our dependent variable, we obtain a p-value of 0.0765. We conclude that an increase in the H1B visa cap increased the likelihood of observing Chinese/Indian inventor names on herbal patents. As a further robustness check, we provide additional placebo tests for our difference in differences (DD) analysis using the text-similarity based control group. We also see slightly larger, yet significant results for our DD analysis. Results are available with authors.

Inventor educational backgrounds using LinkedIn. Inventor background can also provide information about whether herbal patent inventors are more likely to be first generation migrant inventors. Our herbal patents sample contains 4,854 unique inventors, of which 1,005 unique inventors are of Chinese or Indian ethnicities. We randomly sample 552 inventors from the Chinese/Indian inventor population (55% 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, and 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 percent 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. Tables are included in the appendix (tables A6–A8).

Lee Fleming inventor disambiguation dataset. We use the inventor disambiguated patent dataset of Lai *et al*. (2013) to further validate our results. If first generation migrants are patenting contextual knowledge, we would see foreign inventors moving to the U.S. during the shock period, and subsequently writing herbal patents.

The Lai *et al*. (2013) dataset provides us with a disambiguated set of inventors for patents filed between 1975 and 2010, and therefore we can track the location of inventors over time. We used the patent numbers and related patent numbers to match our data with the disambiguated patent inventor database. The dataset offers two classification schemes, “upper” and “lower”, based on how permissive the inventor disambiguation is. We use both upper and lower schemes, and report results for the upper scheme. In total, 2,267 patents and 4,861 inventors were matched, for a resulting dataset of 45,213 inventor-patent pairs. This dataset thus contains the entire patenting history of the 4,861 inventors between the years of 1977–2010, allowing us to track how the patenting behavior of inventors changes over time. Of the 4,861 inventors, we only observe changes in country level location for 53 inventors, limiting our ability to track location of inventors over time.[17]

Patent matching. In addition to our patent matching based on textual similarity of medicinal application area, we tested our specifications using the matching method proposed by Jaffe *et al*. (1993). The main drawback of this method is that by the nature of our dataset, many of our patents are from the same patent class. Therefore, it may not control for our main potential confounding variable, what kind of treatments the patent is intended for. As in Jaffe *et al*. (1993), for each herbal patent, we collected a control patent from the same 3-digit IPC class, in the same year, and with the closest application date. Using the same regression specifications, we obtain robust results, which are available with the authors.

[17] Direct measurements of whether inventors are first-generation migrants using this dataset are difficult to implement. To measure migration, we need inventors that 1) apply for patents abroad prior to our visa shock period, 2) move to the US, and 3) apply for additional patents using their updated address during the shock. This approach is limited by the small number of inventors in our dataset that document changes in location. Nonetheless, this robustness check provides further validation of our results.

Clustering. In our analysis related to Hypothesis 1, we have clustered at the patent assignee level. This is because we believe that the error terms will be correlated for patents with the same assignee. For instance, company specific HR policies may affect the proportion of ethnic Chinese/Indian inventor names on herbal patents filed by the company. An alternative, broader level at which to cluster at would be the IPC class of the patents. However, this reduces the effective number of clusters to <40, which may be too small for unbalanced panels (Cameron and Miller, 2015). Therefore, we believe that the assignee level is an appropriate level to cluster standard errors.

Difference-in-difference specifications. As an alternative to the triple differences specification, we also run separate difference-in-difference (DD) specifications for the cap-subject and cap-exempt groups. We would expect there to be a significant effect of the visa shock for the cap-subject groups, but not for the cap-exempt groups. Indeed, this is the case, and we report the results under these specifications in table A9 of the appendix.

Nonlinear specifications. We also test nonlinear specifications of equation (1) and present results in the appendix (Table A10). The dependent variables in our data contain many zeroes, and therefore nonlinear specifications may better fit the data, and we would not have to worry about predictions that are out of range. We consider three separate models: Poisson, logistic regression, and conditional logit models. We estimated all three nonlinear models using the same specification as (1). As expected from the results in Table 5, we see a positive and significant effect of the H1B visa shock on patents by firms affected by the visa cap. Furthermore, we see the standard errors decrease as the model fit increases, suggesting that nonlinear models further support our hypotheses. It should be noted that treatment effect in nonlinear differences in differences model has the same sign as the interaction term (Puhani, 2012), allowing us to interpret the sign and significance directly.[18]

Discussion and Conclusion

We studied the role of ethnic migrant inventors in transferring contextual knowledge across borders and exploit an exogenous shock to immigration and a list of patenting entities excluded from this shock to present robust econometric results. Our triple differences results show that there is more than a twofold

[18] When using nonlinear models such as Logit or Probit to identify treatment effects, the common trends assumption is violated and therefore the coefficient on the interaction term does not correspond to the actual treatment effect (Athey and Imbens, 2006). It has been shown that generally the interaction term in nonlinear models need not have the same magnitude, and that it may have different signs from the marginal effects of the interaction term (Ai and Norton, 2003). However, it can be shown that the coefficient on the interaction term in a nonlinear differences in differences model does have the same sign as the treatment effect (Puhani, 2012). We find in our case that indeed, nonlinear models have the same sign as our DD and DDD models, and that the statistical significance is also preserved.

increase in the likelihood of observing ethnic Chinese/Indian inventor names on herbal patents for the cap-subject sample, after the treatment. Our results are robust to controlling for citation counts, using different control groups, and to serial correlation. We also see that herbal patents with more contextual knowledge are more likely to have Chinese/Indian inventor names. A one standard deviation increase in contextual knowledge in a patent (measured by Google N-Grams) is associated with a 26 percent increase in the likelihood of having Chinese/Indian names on an herbal patent. Furthermore, during periods of high immigration, this association increases so that high contextual knowledge patents are 35 percent more likely to be filed by Chinese or Indian inventors, suggesting migrant inventors play a key role in transmitting contextual knowledge. Also, having any Chinese/Indian inventors on a patent is associated with a 56 percent and 82 percent greater chance of being cited by other Chinese/Indian inventors for control and herbal patents, respectively, compared to a control patent filed by non-Chinese/Indian inventors filed within the first month of publication of the focal patent. One year after publication, the probability of Chinese/Indian inventors citing an herbal patent by other Chinese/Indian inventors decreases by about 2 percentage points, indicating that non-Chinese/Indian inventors are more likely to cite such patents over time Finally, we find that during the visa shock period, the probability of knowledge recombination for patents with Chinese/Indian inventor names decreases by 5 percentage points for cap-subject herbal patents. In summary, our results support our hypotheses that contextual knowledge is codified by ethnic migrant inventors, spread through ethnic networks, but recombined by inventors belonging to non-ethnic (i.e. non-Chinese/Indian) backgrounds.

Importance of herbal remedies to western bio-pharma industry and western science

An important question here is how central herbal remedies are to the western bio-pharma industry and western science in general. Here we present a few stylized facts. In 2016, the herbal remedies market in the U.S. was worth $5.4 billion dollars and is forecasted to grow to $6.6 billion dollars by 2021 (Mintel, 2016). Examples of products launched in this segment include Metamucil (Proctor and Gamble), Benefiber (GlaxoSmithKline) and Fibercon (Pfizer), among others (Euromonitor, 2016). Table A11 in the appendix lists companies and their market shares in the herbal product market.

Within broader western scientific research, herbal and natural ingredients have been cited as key sources for drug discovery (Doak et. al. 2014), and prior literature documents that between 1981 and 2014, at least 33 percent of all new chemical entities (NCEs) introduced were natural product derived (Newman and Cragg, 2007). This is further reflected in the secondary data we gathered from PubMed. We utilized the PubMed Dietary Supplements Subset and searches using 499 of our herbs resulted in 658,488 articles on PubMed, published in 11,974 unique scientific journals. Figure 5 plots the number of articles published about herbal remedies over time for all journals, and also for selective journals such as *Science*, *Nature* and the *New England Journal of Medicine*.

[Insert Figure 5 Here]

Contributions of our study

Our results contribute to several literatures, including the literature on inventor mobility and knowledge flows, the role of context in innovation, skilled migration, and the microfoundations of knowledge recombination. By outlining *contextual knowledge* as knowledge embedded in its cultural, religious, and linguistic context, we arguably connect the previously disconnected literature of knowledge flows and the literature on cross-national variation in context and fills an important white space in the innovation literature. Similar to Jensen and Szulanski (2004), who argued that institutional distance increases stickiness of knowledge and impedes its transfer, we argue that the cultural and linguistic context affects knowledge flows across borders and contextual knowledge will be codified and transferred by migrant inventors who were previously embedded in the home context. Our insights contribute to the recent call in the strategy literature for firms to develop contextual intelligence (Khanna, 2015, Dhanaraj and Khanna, 2011) and suggests that hiring ethnic migrants might lay the microfoundations to building contextual intelligence.

Our results contribute to the literature on skilled migration and Diaspora. Recent research and the policy debate in this literature (Kerr and Lincoln, 2010, Kerr *et al*., 2012, Doran *et al*., 2016) has focused on the job creation effects of the H1-B program.[19] We side-step that debate in the literature but highlight the role that migrant inventors can play in transferring contextual knowledge across borders. Our results are related to the results reported by researchers studying the impact of immigration of Russian mathematicians into the U.S., post-collapse of the Soviet Union. Borjas and Doran (2012) showed that Russian mathematicians were relatively advanced compared to the west in mathematical fields such as partial differential equations, operator theory, and symplectic topology. Ganguli (2015) builds on their data and documents that citations to Soviet-era work increased significantly with the arrival of immigrants. In a new working paper, Agrawal *et al*. (2013) show that collaboration rose disproportionately in Soviet rich relative to Soviet poor fields after 1990. Moser *et al*. (2014) have a similar finding related to the migration of German Jewish migrants into the U.S. from Nazi Germany and show that migrants encouraged innovation by attracting new American inventors to their fields. In the broader population, Hunt and Gauthier-Loiselle (2010) document that a one percentage point increase in immigrant college graduates' population share increases patents per capita by 9–18 percent. Our finding that there is a slow diffusion of contextual knowledge to non-Chinese/Indian inventors who employ recombination is also in the spirit of

[19] Kerr and Lincoln (2010) find that changes in H1-B admission levels influence the rate of Indian and Chinese patenting in cities and firms dependent upon the program relative to their peers. Kerr *et al*. (2012) finds overall employment of skilled workers to be related to increased skilled immigrant employment by the firm. However Doran *et al*. (2016) find that H1-B visas crowd out firms' employment of other workers.

"national learning by immigration" (i.e. unintended knowledge flows that result from cross-border migration, and that often accrue to firms other than the hiring firms) documented by Oettl and Agrawal (2008). Our results are also relevant to the expanding literature on the role of Diaspora in facilitating knowledge transfer and innovation outcomes across borders (Nanda and Khanna, 2010, Hernandez, 2014, Choudhury 2015, etc.).

Our results also contribute to the broader strategy literature on the micro-foundations of knowledge recombination (Allen, 1977, Fleming 2001) and heeds the call to illuminate the micro-foundations of innovation within firms (e.g., Felin and Foss, 2005). The recent literature in this area includes Gruber *et al*. (2013) who have studied how individual characteristics of inventors (e.g. their educational background and whether or not they are scientists versus engineers) affect the breadth of their technological recombinations. Other recent papers (Fleming *et al*., 2007, Paruchuri and Awate, 2016) in this literature study the characteristics of individual inventor network positions on their ability to engage in recombination.[20] Our findings contribute to this literature and suggest that while ethnic migrant inventors might transfer the novel contextual knowledge into the boundary of the western firm, recombination is likely to be done by the non-ethnic inventor. This indicates a complementary relationship between the ethnic migrant inventor and the non-ethnic inventor from the perspective of knowledge recombination, an insight that is related to the literature on concurrent sourcing of complementary components for knowledge recombination (Parmigiani and Mitchell, 2009; Hess and Rothaermel, 2011).[21] Our results are especially related to Hess and Rothaermel (2011), who build on Arora and Gambardella (1990) and argue that star scientists act as bridges linking the firm to complementary, non-redundant knowledge in other organizations. Our insights also contribute to the broader literature on intra-firm knowledge recombination (Carnabuci and Operti, 2013, Karim and Kaul, 2014).

Limitations and directions for future research

Our study has various limitations, one of which is external validity. We have studied contextual knowledge with respect only to herbal patents filed in the U.S. Other types of contextual knowledge may exhibit different patterns of transfer and recombination, and results might vary across countries as well. Our dependent variables are also limited in that they are merely proxy for the extent of Chinese/Indian inventor activity, though we try to address this issue using LinkedIn data in the robustness checks. Finally, by the nature of our natural experiment, we are capturing the effects of immigration through the marginal

[20] While Fleming *et al*. (2007) study brokerage positions of individual inventors, Paruchuri and Awate (2016) study the reach of inventors in the intra-firm network and their span of structural holes. Other papers in this literature include Nerkar and Paruchuri, 2005, Audia and Goncalo, 2007 and Tzabbar, 2009.

[21] The strategy literature has shown that firms increasingly rely on a combination of internal and external knowledge sourcing for purposes of recombination. Parmigiani and Mitchell (2009) argue that concurrent sourcing of complementary knowledge increases with greater within-firm shared expertise.

H1B visa candidate, a highly skilled individual. More general increases in immigration may have different impacts on contextual knowledge flows and recombination.

Future research can explore specific dimensions of context such as language, religion, cultural attributes and institutional factors that influence the production of contextual knowledge and can explore the role of ethnic scientists in the production and spread of such knowledge. Future work could also focus on studying whether ethnic migrants are responsible for codifying contextual knowledge in forms other than patenting, i.e. in forms such as academic publications and business practices. Another potential area of research is to study conditions under which contextual knowledge from non-western settings recombines with existing western knowledge and outcomes of such recombination. The broader goal of research on contextual knowledge should aim at understanding when and why contextual knowledge is important to transfer and recombine, and how firms augmenting contextual knowledge to the knowledge production function could lead to the appropriation of strategic rents. In the broader strategy literature, scholars could also study the role of skilled ethnic migrants in transferring knowledge underlying cultural goods and services, across borders.[22] Also, we provide evidence of recombination of contextual knowledge (after its transfer by migrants), by inventors belonging to other ethnicities: our results are suggestive of a possible complementary relationship between ethnic migrant inventors (who introduce novel contextual knowledge to the firm) and non-ethnic inventors (who recombine such knowledge, in the spirit of recombinant creation). Future research could explore whether or not there is a more general complementary relationship and overlap in intrafirm social network between newly hired mobile inventors ("knowledge introducers"?) and existing inventors engaged in knowledge recombination ("recombinants"?). Future research could explore how other inventor characteristics (beyond their ethnicity) affect the likelihood that an inventor will engage in recombination of contextual knowledge.

In conclusion, our research introduces a novel categorization of knowledge based on the context in which such knowledge is embedded and identifies a mechanism, i.e. the migration of skilled inventors across borders, in identifying how such knowledge is transferred across borders. Our research also sheds light on an important mechanism (inventor mobility and skilled migration) related to the micro-foundations of intra-firm knowledge recombination and suggests that there is a likely complementary relationship between migrant scientists and other scientists, from the perspective of knowledge recombination. Our results have managerial implications for firms engaged in R&D and cross-border sourcing of ideas and policy implications for the current policy debate around high-skilled immigration and the effectiveness of temporary worker programs such as the H1B.[23]

[22] There is a rich literature in strategy on cultural goods (e.g. Lampel *et al*., 2000) but lack of empirical work linking migration of ethnic knowledge workers and the spread of cultural goods across borders.

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

References

Agrawal, Ajay, Cockburn, I., McHale, J. (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. Journal of Economic Geography, 6(5): 571-591.

Agrawal, Ajay, Avi Goldfarb, and Florenta Teodoridis. Does Knowledge Accumulation Increase the Returns to Collaboration?. No. w19694. National Bureau of Economic Research, 2013.

Agrawal, Ajay, Kapur, D., McHale, J. (2008) How do spatial and social proximity influence knowledge flows? Evidence from patent data. Journal of Urban Economics 64(2): 258-269.

Agrawal, Ajay, 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, Gautam, and Curba Morris Lampert. "Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions." Strategic management journal 22.6-7 (2001): 521-543.

Ai, Chunrong, and Edward C. Norton. "Interaction terms in logit and probit models." Economics letters 80.1 (2003): 123-129.

Alcacer, Juan, and Michelle Gittelman. "Patent citations as a measure of knowledge flows: The influence of examiner citations." The Review of Economics and Statistics 88.4 (2006): 774-779.

Allen, Thomas J. "Managing the flow of technology: technology transfer and the dissemination of technological information within the R and D organization." (1977).

Almeida, Paul, and Bruce Kogut. "Localization of knowledge and the mobility of engineers in regional networks." Management science 45.7 (1999): 905-917.

Almeida, Paul, Anupama Phene, and Sali Li. "The influence of ethnic community knowledge on Indian inventor innovativeness." Organization Science 26.1 (2014): 198-217.

Ambekar, A., Ward, C., Mohammed, J., Male, S., & Skiena, S. (2009, June). Name-ethnicity classification from open sources. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining (pp. 49-58). ACM.

Arora, Ashish, and Alfonso Gambardella. "Complementarity and external linkages: the strategies of the large firms in biotechnology." The Journal of Industrial Economics (1990): 361-379.

Arrow, Kenneth. "Economic welfare and the allocation of resources for invention." The rate and direction of inventive activity: Economic and social factors. Princeton University Press, 1962. 609-626.

Athey, Susan, and Guido W. Imbens. "Identification and inference in nonlinear difference-in-differences models." Econometrica 74.2 (2006): 431-497.

Berry, Heather, Mauro F. Guillén, and Nan Zhou. "An institutional approach to cross-national distance." Journal of International Business Studies 41.9 (2010): 1460-1480.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How Much Should We Trust Differences-in-Differences Estimates?" The Quarterly Journal of Economics 119.1 (2004): 249-275.

Borjas, George J., and Kirk B. Doran. "The Collapse of the Soviet Union and the Productivity of American Mathematicians." The Quarterly Journal of Economics (2012): qjs015.

Breschi, Stefano, and Francesco Lissoni. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." Journal of Economic Geography (2009): lbp008.

Brézillon, Patrick. "Context in problem solving: a survey." The Knowledge Engineering Review 14.01 (1999): 47-80.

Cameron, A. Colin, and Douglas L. Miller. "A practitioner's guide to cluster-robust inference." Journal of Human Resources 50.2 (2015): 317-372.

Carnabuci, Gianluca, and Elisa Operti. "Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination." Strategic Management Journal 34.13 (2013): 1591-1613.

Chetty, Raj, Adam Looney, and Kory Kroft. "Salience and Taxation: Theory and Evidence." The American Economic Review 99.4 (2009): 1145.

Choudhury, Prithwiraj. "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 (2015): lbv025.

Cohen, Wesley M., and Daniel A. Levinthal. "Absorptive capacity: A new perspective on learning and innovation." Administrative science quarterly (1990): 128-152.
Cowan, Robin, and Dominique Foray. "The economics of codification and the diffusion of knowledge." Industrial and corporate change 6.3 (1997): 595-622.
Doran, Kirk, Gelber, Alexander, and Isen, Adam, "The Effect of High-Skill Immigration Policy on Firms: Evidence from H-1B Visa Lotteries", Working paper (2016).
Dasgupta, Partha, and Paul A. David. "Toward a new economics of science." Research policy 23.5 (1994): 487-521.
Dhanaraj, Charles, and Tarun Khanna. "Transforming mental models on emerging markets." Academy of Management Learning & Education 10.4 (2011): 684-701.
Doak, Bradley Croy, *et al*. "Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates." Chemistry & biology 21.9 (2014): 1115-1142.
Docquier, F, Rapoport, H. (2012) Globalization, brain drain, and development. Journal of Economic Literature 50.3: 681-730.
Euromonitor. (2016) "Herbal/Traditional Products in the US" Retrieved March 3, 2017 from Euromonitor Passport database.
Felin T, Foss NJ. 2005. Strategic organization: A field in search of micro-foundations. Strategic Organization 3(4): 441.
Fleming, Lee. "Recombinant uncertainty in technological search." Management science 47.1 (2001): 117-132.
Fleming, Lee, Santiago Mingo, and David Chen. "Collaborative brokerage, generative creativity, and creative success." Administrative science quarterly 52.3 (2007): 443-475.
Foley, C. F, Kerr, W.R. (2013) Ethnic innovation and US multinational firm activity.' Management Science 59.7: 1529-1544.
Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. "The mover's advantage: The superior performance of migrant scientists." Economics Letters 122.1 (2014): 89-93.
Ganguli, Ina. "Immigration and Ideas: What Did Russian Scientists "Bring" to the United States?" Journal of Labor Economics 33.S1 Part 2 (2015): S257-S288.
Ghemawat, Pankaj. "Distance still matters." Harvard business review 79.8 (2001): 137-147.
Gruber, Marc, Dietmar Harhoff, and Karin Hoisl. "Knowledge recombination across technological boundaries: scientists vs. engineers." Management Science 59.4 (2013): 837-851.
Hambrick, Donald C., and Ian C. Macmillan. "Efficiency of product R&D in business units: The role of strategic context." Academy of Management Journal 28.3 (1985): 527-547.
Henderson, Rebecca M., and Kim B. Clark. "Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms." Administrative science quarterly (1990): 9-30.
Hernandez, Exequiel. "Finding a home away from home effects of immigrants on firms' foreign location choice and performance." Administrative Science Quarterly 59.1 (2014): 73-108.
Hess, Andrew M., and Frank T. Rothaermel. "When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry." Strategic Management Journal 32.8 (2011): 895-909.
Hofstede, Geert. Culture's consequences: International differences in work-related values. Vol. 5. sage, 1984.
Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. "How much does immigration boost innovation?." American Economic Journal: Macroeconomics 2.2 (2010): 31-56.
Jensen, Robert, and Gabriel Szulanski. "Stickiness and the adaptation of organizational practices in cross-border knowledge transfers." Journal of International Business Studies 35.6 (2004): 508-523.
Karim, Samina, and Aseem Kaul. "Structural recombination and innovation: Unlocking intraorganizational knowledge synergy through structural change." Organization Science 26.2 (2014): 439-455.

Kerr, W. R. (2008) Ethnic scientific communities and international technology diffusion. The Review of Economics and Statistics 90.3: 518-537.
Kerr, William R., and William F. Lincoln. The supply side of innovation: H-1B visa reforms and US ethnic invention. No. w15768. National Bureau of Economic Research, 2010.
Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln. Skilled immigration and the employment structures and innovation rates of US firms. Technical report, 2012.
Khanna, Tarun. "A case for contextual intelligence." Management International Review 55.2 (2015): 181-190.
Kostova, Tatiana. "Success of the transnational transfer of organizational practices within multinational companies." Unpublished doctoral dissertation, University of Minnesota (1996).
Lai, Ronal, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming. "Disambiguation and Co-Authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)." hdl:1902.1/15705, Harvard Dataverse, V5
Lampel, Joseph, Theresa Lant, and Jamal Shamsie. "Balancing act: Learning from organizing practices in cultural industries." Organization science 11.3 (2000): 263-269.
Lin, Xiaohua, and Richard Germain. "Sustaining satisfactory joint venture relationships: The role of conflict resolution strategy." Journal of International Business Studies 29.1 (1998): 179-196.
March, James G. "Exploration and exploitation in organizational learning." Organization science 2.1 (1991): 71-87.
Mintel. (2016) "Homeopathic and Herbal Remedies – US – November 2016" Retrieved March 3, 2017 from Mintel Academic.
Moser, Petra, Alessandra Voena, and Fabian Waldinger. "German Jewish émigrés and US invention." The American Economic Review 104.10 (2014): 3222-3255.
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, Richard R., and Sidney G. Winter. An evolutionary theory of economic change. Harvard University Press, 1982.
Nerkar, Atul, and Srikanth Paruchuri. "Evolution of R&D capabilities: The role of knowledge networks within a firm." Management Science 51.5 (2005): 771-785.
Newman, David J., and Gordon M. Cragg. "Natural Products as Sources of New Drugs over the Last 25 Years ⊥." Journal of natural products 70.3 (2007): 461-477.
Oettl, Alexander, and Ajay Agrawal. "International labor mobility and knowledge flow externalities." Journal of International Business Studies 39.8 (2008): 1242-1260.
Parmigiani, Anne, and Will Mitchell. "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 (2009): 1065-1091.
Paruchuri, Srikanth, and Snehal Awate. "Organizational knowledge networks and local search: The role of intra-organizational inventor networks." Strategic Management Journal (2016).
Polanyi, Michael, 1966, The Tacit Dimension (Routledge & Kegan Paul, London).
Porter, M. 1990. The Competitive Advantage of Nations. Free Press, New York
Puhani, Patrick A. "The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models." Economics Letters 115.1 (2012): 85-87.
Rosenkopf, Lori, and Paul Almeida. "Overcoming local search through alliances and mobility." Management science 49.6 (2003): 751-766.
Saxenian, A. (1999) Silicon Valley's new immigrant entrepreneurs. Vol. 32. San Francisco: Public Policy Institute of California.
Schumpeter, Joseph Alois. Business cycles. Vol. 1. New York: McGraw-Hill, 1939.
Singh, Jasjit, and Matt Marx. "Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity." Management Science 59.9 (2013): 2056-2078.
Song, Jaeyong, Paul Almeida, and Geraldine Wu. "Learning–by–Hiring: When is mobility more likely to facilitate interfirm knowledge transfer?." Management science 49.4 (2003): 351-365.

Szulanski, Gabriel. "Exploring internal stickiness: Impediments to the transfer of best practice within the firm." Strategic management journal 17.S2 (1996): 27-43.
Thompson, Peter, and Melanie Fox-Kean. "Patent citations and the geography of knowledge spillovers: A reassessment." American Economic Review (2005): 450-460.
Tsang, Eric WK, and Paul SL Yip. "Economic distance and the survival of foreign direct investments." Academy of Management Journal 50.5 (2007): 1156-1168.
Von Hippel, Eric. "“Sticky information” and the locus of problem solving: implications for innovation." Management science 40.4 (1994): 429-439.
Zhou, Kevin Zheng, and Fang Wu. "Technological capability, strategic flexibility, and product innovation." Strategic Management Journal 31.5 (2010): 547-561.

Table 1. *Herb Names and Frequencies*

Herb Name	Patents with Herb	Percent	Cumulative
soybean	356	3.06	3.06
Soy	279	2.4	5.45
Aloe	257	2.21	7.66
grape	253	2.17	9.83
Green Tea	234	2.01	11.84
ginseng	182	1.56	13.4
rosemary	166	1.43	14.83
cocoa	156	1.34	16.17
licorice	151	1.3	17.46
jojoba	148	1.27	18.73

Notes - Table 1 shows the 10 most frequent herbs within our dataset of herbal patents. Single patents may contain more than one herb name. Percentages across all patent-herb pairs.

Table 2. *Cap-Exempt Assignees and Patent Counts*

Assignee	Patent Counts	Assignee	Patent Counts
The Regents Of The University Of California	14	Phytomyco Research Corporation	5
Bristol-Myers Squibb Company	13	Unigen Pharmaceuticals, Inc.	5
Rutgers, The State University Of New Jersey	12	Amgen Inc.	5
Board Of Trustees Of Michigan State University	11	Pioneer Hi-Bred International, Inc.	5
Genentech, Inc.	11	Univera Pharmaceuticals, Inc.	4
Merck & Co., Inc.	10	Regeneron Pharmaceuticals, Inc.	4
The Trustees Of Columbia University In The City Of New York	7	Regents Of The University Of Minnesota	4
University Of Tennessee Research Foundation	6	Board Of Regents, The University Of Texas System	4

Notes - Table 2 displays the most frequent cap-exempt assignees in our dataset. Included in the list are a number of for-profit firms. H1B visa regulations on cap-exemption state the employee must be hired to work "at" universities or nonprofits, not "by" those employers.

Table 3. *Summary Statistics of Control and Herbal Patents*

	Control Patents		Herbal Patents	
	(1)	(2)	(3)	(4)
	Exempt	Capped	Exempt	Capped
	mean/sd	mean/sd	mean/sd	mean/sd
Has Chinese/Indian	0.234	0.185	0.333	0.287
(ETHNIC INVENTOR)	(0.425)	(0.388)	(0.473)	(0.453)
Has European	0.964	0.878	0.904	0.819
	(0.188)	(0.328)	(0.296)	(0.385)
Chinese/Indian Count	0.277	0.337	0.481	0.689
	(0.552)	(1.018)	(0.771)	(1.465)
European Count	2.175	1.929	1.926	1.872
	(1.254)	(1.616)	(1.364)	(1.735)
Fraction Chinese/Indian	0.096	0.103	0.180	0.172
	(0.201)	(0.251)	(0.297)	(0.309)
Fraction European	0.855	0.796	0.766	0.674
	(0.257)	(0.354)	(0.333)	(0.407)
Citations Count	29.380	19.027	6.785	7.967
	(52.725)	(25.573)	(8.747)	(14.072)
Inventor Count	2.606	2.586	2.556	3.133
	(1.291)	(2.043)	(1.505)	(2.370)
Observations	137	1923	135	1925

Notes - Table 3 presents summary statistics for inventor ethnicities and citations for control patents (columns 1-2) and herbal patents (columns 3-4). We present our main dependent variable *(ETHNIC INVENTOR)* defined earlier as whether a patent has any inventors with Chinese/Indian names (row 1), a dummy for whether a patent has any inventors with European names (row 2), the number of inventors on a patent by ethnicity (rows 3-4), and the fraction of inventors of a certain ethnicity (rows 5-6). "European" is a term for various Western ethnicities. We see that herbal patents are more likely to have Chinese/Indian inventors, and have almost twice as many Chinese/Indian inventors as control patents. Similarly, the fraction of Chinese/Indian inventors is greater for herbal patents. Herbal patents have fewer citations. Within herbal and control patent groups, we report summary statistics for the cap-exempt sub-samples in columns 1 and 3 respectively and summary statistics for the cap-subject sub-samples in columns 2 and 4 respectively. Results of T-tests are reported in Table 4.

Table 4. *T-Test Results for Control and Herbal Patents, on Average*

	(1) Control Patents mean/sd	(2) Herbal Patents mean/sd	(3) Difference b/se
Has Chinese/Indian	0.188	0.290	-0.102***
(ETHNIC INVENTOR)	(0.391)	(0.454)	(0.013)
Has European	0.883	0.824	0.059***
	(0.321)	(0.381)	(0.011)
Chinese/Indian Count	0.333	0.676	-0.343***
	(0.994)	(1.430)	(0.038)
European Count	1.946	1.875	0.070
	(1.595)	(1.713)	(0.052)
Fraction Chinese/Indian	0.102	0.172	-0.070***
	(0.248)	(0.308)	(0.009)
Fraction European	0.800	0.680	0.120***
	(0.349)	(0.403)	(0.012)
Citations Count	19.715	7.890	11.825***
	(28.297)	(13.787)	(0.694)
Inventor Count	2.587	3.095	-0.508***
	(2.002)	(2.327)	(0.068)
Cap-subject assignee	0.933	0.934	-0.001
	(0.249)	(0.248)	(0.008)
Observations	2060	2060	4120

Notes - Table 4 presents t-tests for control and herbal patents, on average, across both cap-exempt and cap-subject sub-samples. Column (1) and (2) present the means and standard deviations for the respective population groups. Column (3) presents the differences and standard errors for t-statistics. Our main dependent variable, *Has Chinese/Indian (ETHNIC INVENTOR)*, and *Has European* are indicators for whether a patent has any authors of said categories. *Chinese/Indian Count*, *European Count* report the number of inventors of said categories in each patent. *Fraction Chinese/Indian*, *Fraction European* present the fraction of inventors in said categories. We see herbal patents are significantly more likely to have ethnic inventors compared to the control group, regardless of the measure. There is no statistically significant difference in the number of cap-subject patents or the number of European inventors on a patent across control and herbal patents.

Table 5. *Testing Hypothesis 1 - Triple Difference Estimates*

	(1)	(2)	(3)	(4)
Dependent Variable:	*ETHNIC INVENTOR*	*ETHNIC INVENTOR*	*ETHNIC INVENTOR*	*ETHNIC INVENTOR*
HERB (β_1)	0.154*	0.144*	0.147*	0.148*
	(0.0805)	(0.0808)	(0.0811)	(0.0814)
CAP (β_2)	-0.0302	-0.0431	-0.0425	-0.0410
	(0.0444)	(0.0439)	(0.0439)	(0.0418)
TREAT (β_3)	0.0494	0.243***	0.241***	0.492***
	(0.0718)	(0.0745)	(0.0746)	(0.0753)
HERB x TREAT	-0.142	-0.133	-0.132	-0.117
(γ_1)	(0.121)	(0.121)	(0.121)	(0.114)
CAPx TREAT	-0.0522	-0.0431	-0.0414	-0.0305
(γ_2)	(0.0747)	(0.0741)	(0.0741)	(0.0698)
HERB x CAP	-0.0900	-0.0790	-0.0800	-0.117
(γ_3)	(0.0835)	(0.0838)	(0.0840)	(0.0833)
DDD (δ)	0.236*	0.227*	0.225*	0.213*
	(0.129)	(0.129)	(0.129)	(0.120)
Constant	0.216***	0.0104	0.00606	-0.423***
	(0.0427)	(0.0438)	(0.0444)	(0.0614)
Time Fixed Effects	No	Yes	Yes	Yes
Controls	No	No	Citation Count	Citations & Inventor Count
Observations	4120	4120	4120	4120
Adjusted R^2	0.019	0.031	0.031	0.134

Notes - Table 5 presents results from testing Hypothesis 1 (i.e. contextual knowledge is more likely to be codified by ethnic migrant inventors), estimating equation (1) using OLS. Column (1) reports coefficients for the regression without any controls or fixed effects. Columns (2–4) report coefficients for the same specification, gradually adding controls. There is a significant, time-invariant difference between herbal and control patents in terms of their likelihood of having an ethnic Chinese/Indian inventor(β_1), echoing our t-test results. Furthermore, there is an increase in Chinese/Indian inventors filing herbal patents during the visa shock treatment (β_3). The baseline model in column 1 shows that the visa shock caused a 0.236 increase in the probability of having an ethnic Chinese/Indian inventor on a patent (δ), with a baseline probability of 0.216 (i.e. constant term in column 1). We see that this is in excess of a twofold increase in the likelihood of observing ethnic Chinese/Indian inventors on herbal patents for the cap-subject sample, after the treatment. Adding controls decreases the effect size to 0.213, but the effect is still significant. Cluster robust standard errors in parentheses, clustered at the assignee level.

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

Table 6. *Contextual Knowledge Contained Within Herbal Patent and Likelihood of Observing Chinese/Indian Inventor Names on Patent*

	(1)	(2)	(3)
Dependent Variable:	*ETHNIC INVENTOR*	*ETHNIC INVENTOR*	*ETHNIC INVENTOR*
$CONTEXT$ (β_1)	0.0306***	0.0327***	0.0317***
	(0.00680)	(0.00682)	(0.00621)
Constant	0.300***	-0.140***	-0.588***
	(0.0159)	(0.0292)	(0.0362)
Time Fixed Effects	No	Yes	Yes
Controls	No	No	Citations & Inventor Count
Observations	2039	2039	2039
Adjusted R^2	0.016	0.034	0.134

Notes - Table 6 presents estimation results for equation (5) and studies the association between the extent of contextual knowledge contained within an herbal patent and the likelihood of observing Chinese/Indian inventor names on the patent. Compared to the baseline likelihood of observing Chinese/Indian inventor names on a patent (the constant term of 0.30 in column 1), a one standard deviation increase in contextual knowledge, i.e. a one standard deviation increase in the value of the variable *CONTEXT* (which is equal to 2.58, summary statistics of *CONTEXT* available with authors) increases the likelihood of observing an ethnic Chinese/Indian inventor name on the patent by 26% (using point estimate of β_1 in column 1). The effect is robust for controlling for time fixed effects (Column 2 and 3), and the number of inventors and citations (Column 3). Cluster robust standard errors in parentheses, at the assignee level. The sample size of herbal patents in this analysis is 2039 patents, less than the 2060 patents in our sample because 49 out of the 499 herbs do not have N-Gram values in Google. Dependent variable is an indicator for whether the patent contains Chinese/Indian inventors. *CONTEXT* measures the normalized average inverse log frequencies of all herbs in a patent so that an herb with average contextual information has zero value for the variable *CONTEXT*. There is a positive and significant association between the contextual knowledge in a patent and the likelihood of observing Chinese/Indian inventor names on that patent (β_1), which is consistent with contextual knowledge being more likely to be codified by ethnic migrant inventors.

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

Table 7. *Contextual Knowledge Contained Within Herbal Patent and Likelihood of Observing Chinese/Indian Inventor Names on Patent during Treatment Period*

Dependent Variable	(1) *ETHNIC INVENTOR*	(2) *ETHNIC INVENTOR*	(3) *ETHNIC INVENTOR*
$CONTEXT\ (\beta_1)$	0.0215^{***} (0.00800)	0.0246^{***} (0.00796)	0.0242^{***} (0.00755)
$TREAT\ (\beta_2)$	0.0611^{***} (0.0204)	0.465^{***} (0.0549)	0.710^{***} (0.0526)
$CONTEXT \times TREAT\ (\gamma)$	0.0237^{**} (0.0118)	0.0206^{*} (0.0116)	0.0193^{*} (0.0109)
Constant	0.275^{***} (0.0162)	-0.105^{***} (0.0341)	-0.555^{***} (0.0416)
Time Fixed Effects	No	Yes	Yes
Controls	No	No	Citations & Inventor Count
Observations	2039	2039	2039
Adjusted R^2	0.023	0.035	0.136

Notes - Table 7 presents estimation results for equation (6) and studies the association between the extent of contextual knowledge contained within an herbal patent and the likelihood of observing Chinese/Indian inventor names on the patent during the treatment period. The dependent variable is an indicator for whether the patent contains Chinese/Indian inventor names. *CONTEXT* measures the normalized average inverse log frequencies of all herbs in a patent, so that an herb with average contextual information has zero value for the variable *CONTEXT*. The standard deviation for CONTEXT across all herbs is 2.58. *TREAT* is a dummy for the years 2000-2004. β_1 is the time invariant differences in patent inventor ethnicity for patents with more or less contextual knowledge. β_2 denotes the time-specific difference in patent inventor ethnicity for patents with average context. γ captures the change in the relationship between contextual knowledge and patent inventor ethnicity during the visa shock period. The baseline likelihood of observing Chinese/Indian inventor names on a patent (the constant term of 0.275 in column 1) is the likelihood of observing Chinese/Indian names on a patent with mean values for *CONTEXT* in the non-treatment period. A one standard deviation increase in contextual knowledge, i.e. a one standard deviation increase in the value of the variable *CONTEXT* increases the likelihood of observing an ethnic Chinese/Indian inventor name on the patent by 20% during the non-treatment period (using point estimate of β_1 in column 1). In comparison, the baseline likelihood of observing Chinese/Indian names on a patent with mean values for *CONTEXT* in the treatment period is 0.3361. A one standard deviation increase in contextual knowledge, i.e. a one standard deviation increase in the value of the variable *CONTEXT*, increases the likelihood of observing an ethnic Chinese/Indian inventor name on the patent by 35% during the treatment period (from a baseline of the constant plus β_2 using point estimates of β_1 β_2 and γ in column 1). We see that the treatment period disproportionately increases the association between *CONTEXT* and the likelihood of observing Chinese/Indian inventor names on a patent (from 20% to 35%).Cluster robust standard errors in parentheses, at the assignee level. The sample size of herbal patents in this analysis is 2039 patents, less than the 2060 patents in our sample because 49 out of the 499 herbs do not have N-Grams values in Google. The effect is robust for controlling for time fixed effects (Column 2 and 3), and the number of inventors and citations (Column 3).

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

Table 8. *Testing Hypothesis 2 - Forward Citation Patterns for Herbal and Control Patents*

Dependent Variable:	(1) *FRACTION OF CITATIONS ETHNIC*	(2) *FRACTION OF CITATIONS ETHNIC*	(3) *FRACTION OF CITATIONS ETHNIC*
ETHN (β_1)	0.142***	0.115***	0.106***
	(0.0311)	(0.0305)	(0.0300)
HERB (β_2)	-0.0499*	-0.0426	-0.0102
	(0.0281)	(0.0275)	(0.0290)
TIME (β_3)	0.000301*	0.000549**	0.000383
	(0.000155)	(0.000268)	(0.000282)
$ETHN \times HERB\ (\gamma_1)$	0.116**	0.136***	0.125**
	(0.0514)	(0.0503)	(0.0512)
$HERB \times TIME\ (\gamma_2)$	0.000142	0.000160	0.000155
	(0.000258)	(0.000251)	(0.000280)
$ETHN \times TIME\ (\gamma_3)$	-0.000511**	-0.000356	-0.000301
	(0.000260)	(0.000250)	(0.000248)
$\boldsymbol{ETHN \times HERB \times TIME\ (\delta)}$	-0.000609	-0.000766*	-0.000724*
	(0.000444)	(0.000431)	(0.000433)
Time Fixed Effects	No	Yes	Yes
Controls	No	No	Inventor Count
Constant	0.253***	-20.21***	-14.71***
	(0.0193)	(2.864)	(2.854)
Observations	23963	23560	23560
Adjusted R^2	0.015	0.023	0.087

Notes - Table 8 shows regression results from testing Hypothesis 2 (i.e. contextual knowledge is more likely to spread through ethnic inventor networks), estimating equation (2) using OLS and a dataset of forward citations for herbal patents and control patents. The dependent variable is the fraction of forward citations that have any Chinese/Indian inventor names. $ETHN$ is an indicator for whether the cited patent has Chinese/Indian inventor names. $HERB$ is an indicator for whether the cited patent is herbal, and $TIME$ measures time to citation in months. Column (1) shows the baseline regression and suggests that having any Chinese/Indian inventors on a patent is associated with a 56% and 82% greater chance of being cited by other Chinese/Indian inventors for control and herbal patents respectively, compared to a control patent filed by non-Chinese/Indian inventors filed within the first month of publication of the focal patent. Chinese/Indian citations are slightly increasing over time. Most importantly, an herbal patent with a Chinese/Indian inventor has 17% higher probability of being cited by other Chinese/Indian inventors than similar control patents (i.e. control patents filed by Chinese/Indian inventors). One year after publication, the probability of Chinese/Indian inventors citing an herbal patent by other Chinese/Indian inventors decreases by about 2 percentage points, indicating that non-Chinese/Indian inventors are more likely to cite such patents over time. Cluster robust standard errors, clustered at the level of individual herb names, in parentheses

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

Table 9. *Testing Hypothesis 3 - Arbitrage or Recombination Using Synthetic patent authorship*

	(1)	(2)	(3)	(4)
Sample:	Full Sample	Full Sample	Chinese/Indian authored Patents	Non-Chinese/Indian Patents
Dependent Variable	*RECOMBINED*	*RECOMBINED*	*RECOMBINED*	*RECOMBINED*
TREAT (β_1)	0.191**	0.479***	0.645***	0.380**
	(0.0822)	(0.115)	(0.135)	(0.147)
CAP (β_2)	0.0480	0.0348	0.0346	0.0303
	(0.0490)	(0.0599)	(0.0985)	(0.0791)
CAP x TREAT (γ)	-0.239***	-0.215*	-0.316**	-0.162
	(0.0847)	(0.111)	(0.143)	(0.142)
Time FE	No	Yes	Yes	Yes
Controls	No	Citations & Inventor Count	Citations & Inventor Count	Citations & Inventor Count
Constant	0.235***	-0.0699	-0.228***	-0.0144
	(0.0471)	(0.0683)	(0.0617)	(0.0893)
Observations	2060	2060	598	1462
Adjusted R^2	0.005	0.032	0.033	0.027

Notes - Table 9 shows the effect of the visa shock on knowledge recombination. *RECOMBINED* is an indicator variable for having synthetic compounds in addition to herbs in the patent text, a proxy for knowledge recombination. Columns (1-2) analyze the impact of the visa shock on knowledge recombination using the entire herbal patent dataset. Columns (3-4) report subsample analyses for Chinese/Indian and non-Chinese/Indian patents. Our baseline regression in column 1 shows that recombination through synthetic patenting has increased in the full sample during the shock period (β_1), but there is no significant time-invariant differences between cap-subject and cap-exempt patents (β_2). The point estimate of the interaction term γ is negative and statistically significant across models 1–3, and this suggests that the H1B visa shock decreased the use of synthetic compounds within herbal patents, for the cap-subject sub-sample. Furthermore, as the point estimate of γ in column 3 suggests, the effect seems to be driven by patents with Chinese/Indian inventors. This suggests in the cap-subject sub-sample, during the treatment period, more pure herbal patents (i.e. patents with only herbs and no synthetic compounds) were being filed. We interpret this evidence as suggestive of the fact that first generation migrants were arbitraging knowledge and inventors of other ethnicities were engaged in recombination. Clustered robust standard errors at the assignee level.

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

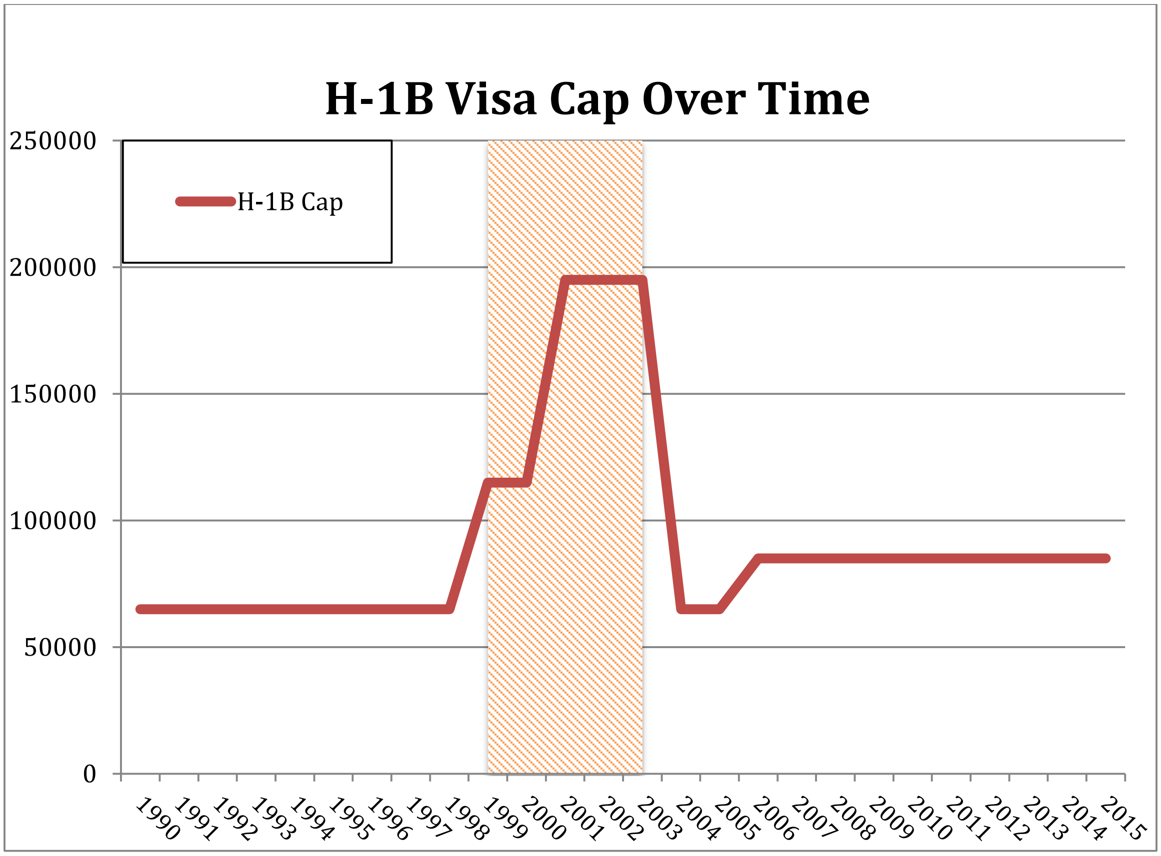

Figure 1. *H1B visa cap over time*

Notes — Figure 1 plots the H1B visa cap over time. The shaded area represents the time for which the American Competitiveness in the 21st Century Act (AC21) was in place. The American Competitiveness and Workforce Improvement Act (ACWIA) was passed in 1998, and the American Competitiveness in the 21st Century Act (AC21) was passed in 2000. As a result of these two legislations, the number of H1B visas increased from 65,000 in 1998 to 115,000 in 1999, up to 195,000 again in 2001, and back down to 65,000 in 2004. The AC21 has a clause that also retroactively increased the quota for 1999 and 2000, past the 115,000 cap set by the ACWIA. After the AC21 Act was passed in 2000, universities and affiliated nonprofits were exempt from the cap, hence the total number of visa issuances can exceed the visa cap. In summary the H1B visa quotas were elevated between 1999 and 2003. In the base case, we consider 2000-2004 as the treatment period (*TREAT*) given that migrants moving to the U.S. would probably need at least a year before they could start filing patents. In robustness checks we relax this constraint.

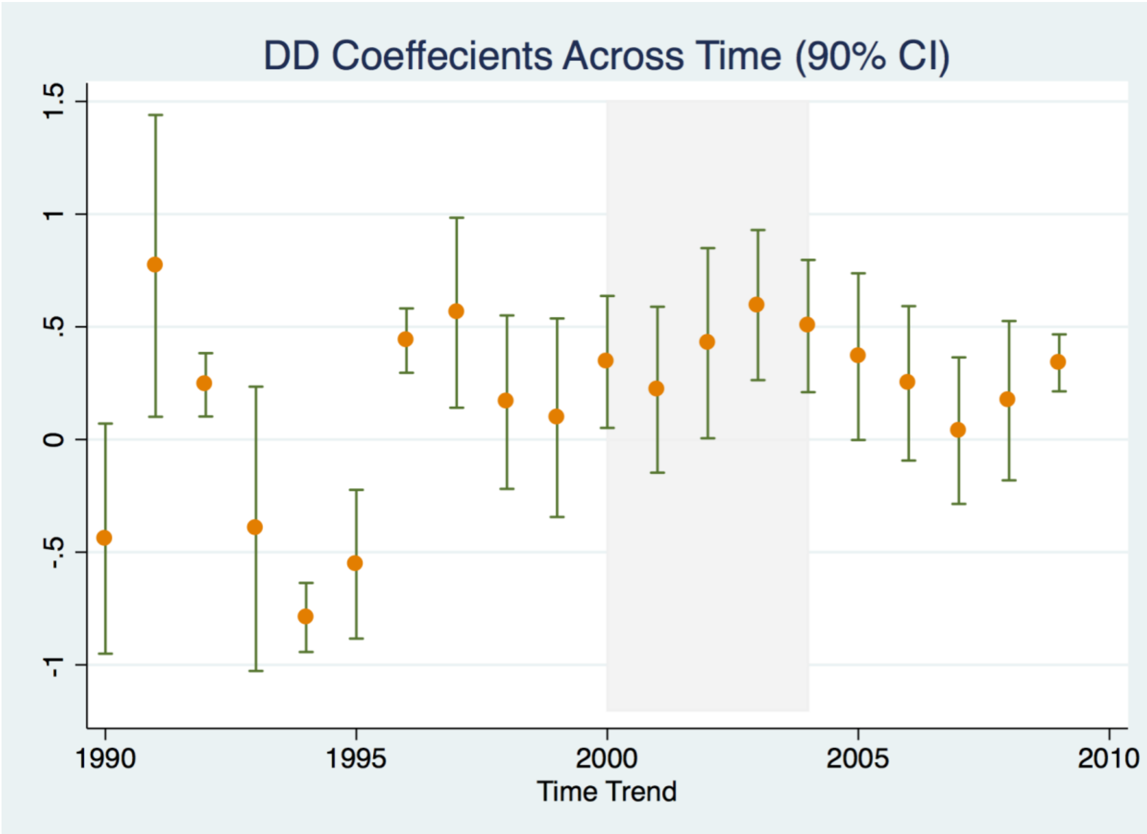

Figure 2. *Effect of Visa Cap on Likelihood of Observing Chinese/Indian Inventors in Herbal Patents over Time*

Notes — Figure 2 plots the effect of the visa cap on the likelihood of observing Chinese/Indian inventors in herbal patents (γ_1) over time. In other words, we plot the mean difference in the likelihood of observing Chinese/Indian inventors in herbal patents across the cap-subject and cap-exempt sub-samples, over time. We plot the DD coefficients along with the 90% confidence intervals. The grey shading represents the time period during which the visa cap was increased. We see that during the treatment period (2000-2004), four out of five coefficients are significantly different from zero, indicating a statistically significant difference in the likelihood of observing Chinese/Indian inventors on herbal patents across cap-subject and cap-exempt sub-samples.

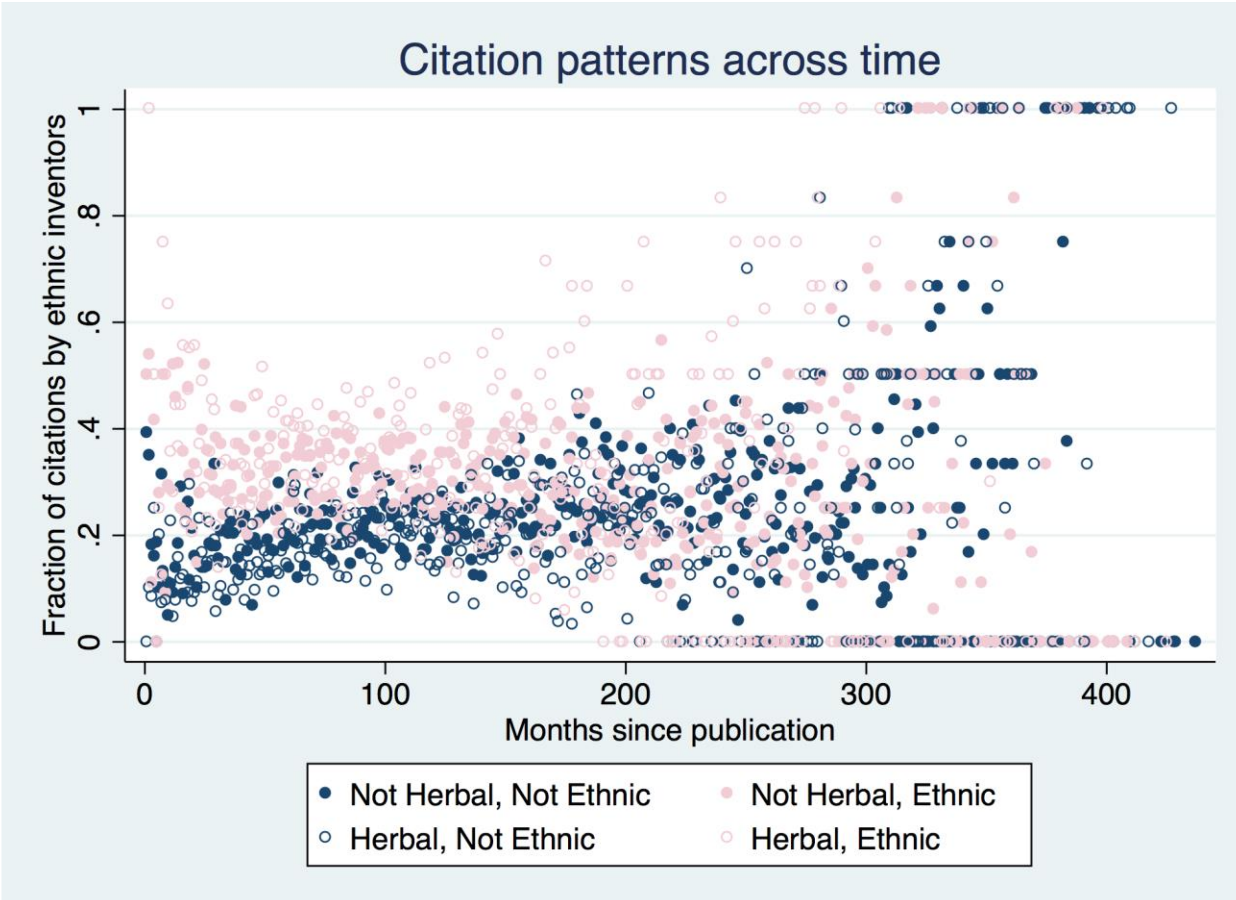

Figure 3. *Citation patterns of herbal and control group patents by Ethnic inventors*

Notes — Figure 3 plots citation patterns over time. Each dot represents a single patent of our herbal and control patent set in a given month, and the fraction of citing patents that have any Chinese/Indian inventor names. Herbal patents are marked with a hollow "o," and control patents with a solid "o." The lighter colored dots represent patents with at least one ethnic inventor, and darker dots are ones without ethnic inventors. Figure 3 reveals a pattern where ethnic inventors cite other ethnic inventors. Furthermore, herbal patents display a divergent pattern, where herbal patents by ethnic inventors are more likely to be cited by other ethnic inventors, but herbal patents by non-ethnic inventors are less likely to be cited by ethnic inventors.

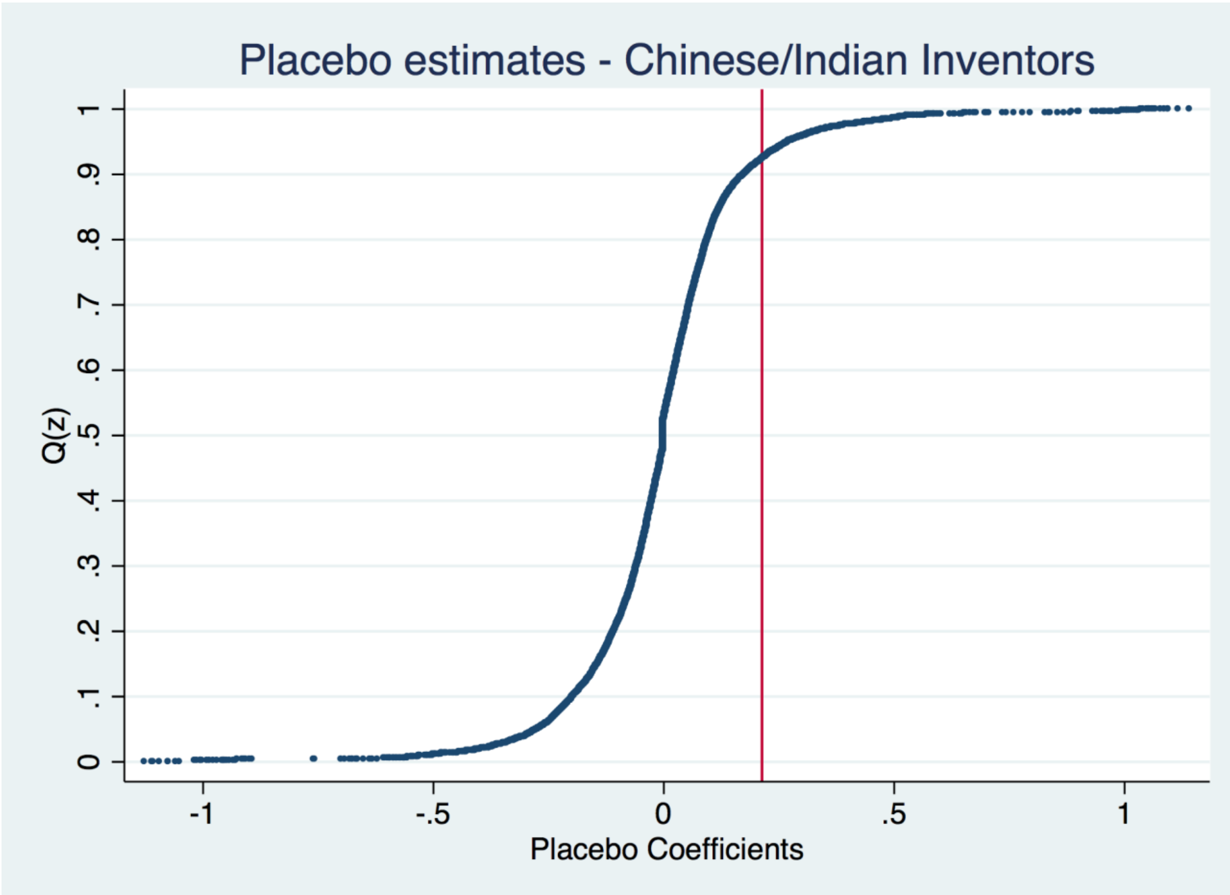

Figure 4. *Placebo test results for DDD analysis*

Notes — Figure 4 plots the results from our placebo test for the DDD analysis. Each blue dot represents a triple differences (δ) coefficient from a randomized placebo triplet. We select the random placebo triplets based on three dimensions – assignment of treatment (done 20 times each), assignment of time period (done 29 times each for the 29 different possible 5-year time periods) and assignment of cap-subject/cap-exempt status (done 20 times each) for a total of $20 \times 20 \times 29 = 11{,}600$random placebo triplets. Since the placebo treatment is randomized within the sample, we should expect to see DDD coefficients as extreme as in Table 5 less than 10% of the time by chance, similar in spirit to a p-value. The vertical red lines denote coefficients from Table 5. Q(z) corresponds to where the vertical line crosses the empirical distribution of coefficients. The Q(z)s for the placebo test, which can be interpreted analogously to p-values, correspond to 0.07647. As Figure 4 indicates, the point estimate for δ that we observed in the fully specified model (column 4) of Table 5 (i.e. 0.213) is likely to be observed less than 10% of the time by chance,

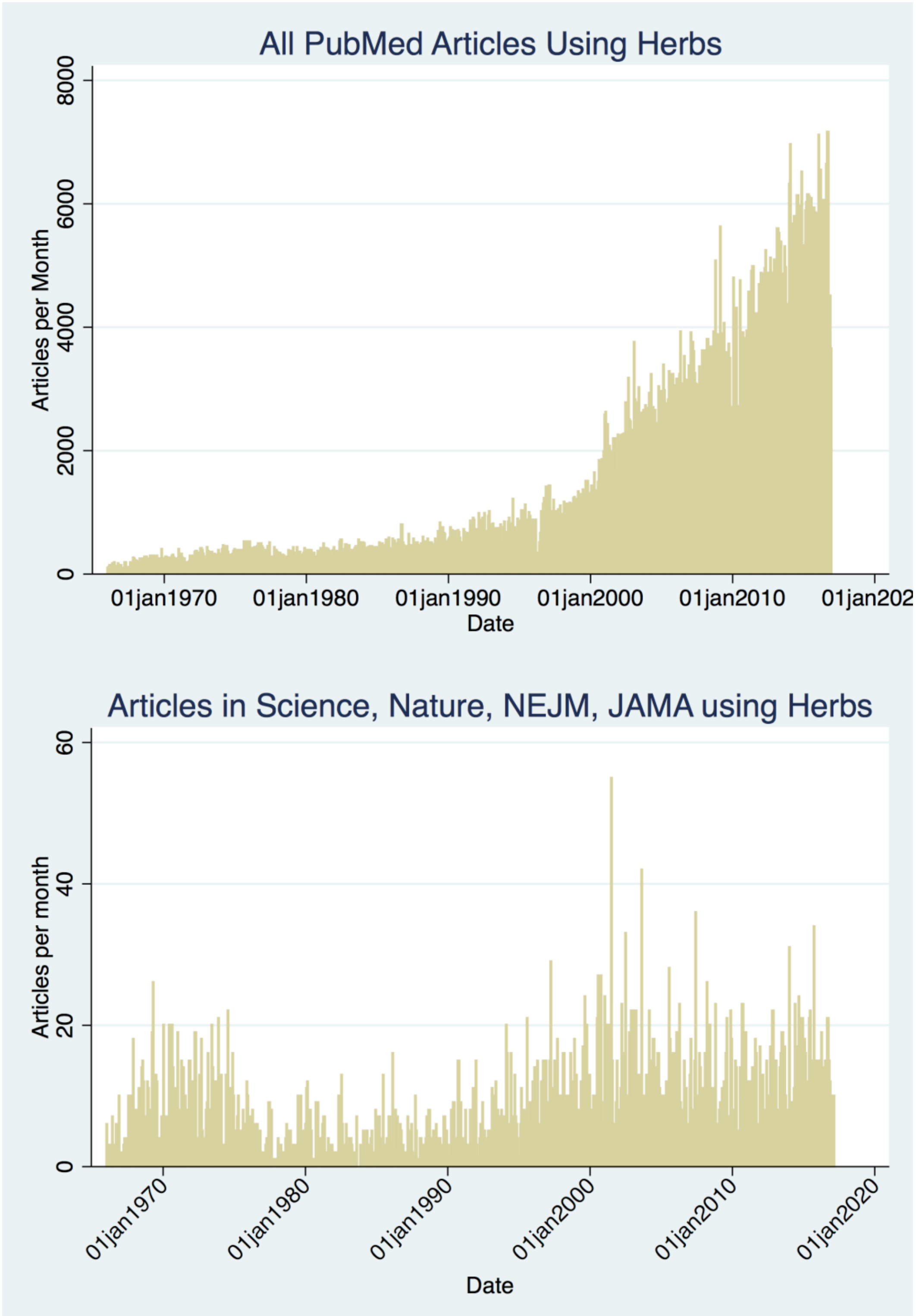

Figure 5. *Articles in western scientific research using herbal remedies*

Notes — Figure 5 plots trends of scientific articles based on herbal remedies on PubMed over time. We see an increase in articles using herbal remedies over time. Restricting the subset to the most impactful journals such as *Science, Nature, the New England Journal of Medicine, etc.*, (bottom panel) also show a general increase in herbal research.

Appendix

Ethnicity robustness checks

In this section, we examine how our definition of ethnicity affects our results. Our ethnicityguesser program offers several different training sets. We use two training sets in particular because they contain categories for Chinese and Indian names. We chose the classifier that had the most extensive training set. Furthermore, we use inventors' full names as our input. Our DD results are robust to which training set we use, and to using inventors' full names. We present simple correlations across the ethnicity measures obtained from each method. The table presents correlations between the number of inventors with the given ethnicity and classification system. Asian is an inclusive term for Chinese, Indian, Korean, Japanese, Vietnamese and Thai. European is an inclusive term for 14 ethnicities in Europe.

Table A1. *Correlations across Ethnicity measures*

	Chinese1 (surname)	Chinese2 (Surname)	Chinese1 (full)	Chinese2 (full)
Chinese1 (surname)	1			
Chinese2 (Surname)	0.8914	1		
Chinese1 (full)	0.881	0.855	1	
Chinese2 (full)	0.815	0.9231	0.8782	1

	Asian1 (surname)	Asian2 (surname)	Asian1 (full)	Asian2 (full)
Asian1 (surname)	1			
Asian2 (surname)	0.9637	1		
Asian1 (full)	0.9503	0.9195	1	
Asian2 (full)	0.9297	0.9394	0.963	1

	Indian1 (surname)	Indian2 (surname)	Indian1 (full)	Indian2 (full)
Indian1 (surname)	1			
Indian2 (surname)	0.9768	1		
Indian1 (full)	0.9279	0.9267	1	
Indian2 (full)	0.9098	0.9172	0.9674	1

	European1 (surname)	European2 (surname)	European1 (full)	European2 (full)
European1 (surname)	1			
European2 (surname)	0.9403	1		
European1 (full)	0.9219	0.9055	1	

European2 (full)	0.8812	0.9297	0.95	1

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. We have bolded out ethnicities we use for our European category. 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 A2. *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

Matching robustness checks

In this section, we examine whether our results are sensitive to how we collect our control groups.

Table A3. *Differences in assignees for herbal and control patents*

Assignee type	Assignee	Fraction of Herbal Patents by Assignee	Fraction of Control Patents by Assignee
Both Herbal and Control	Council of Scientific and Industrial Research	0.045145631	0.005825243
	L'Oreal SA	0.010679612	0.009223301
	Kao Corporation	0.011650485	0.006796117
	The Procter & Gamble Company	0.010194175	0.006796117
	Access Business Group International LLC	0.011650485	0.000970874
Only Herbal	Coty Inc.	0.004368932	-
	Johnson & Johnson Consumer Companies, Inc.	0.004368932	-
	Zenitech, LLC	0.003398058	-
	Laboratoires Expanscience S.A.	0.003398058	-
	Vitacost.com, Inc.	0.002912621	-
Only Control	SmithKline Beecham Limited	-	0.002912621
	Sunovion Pharmaceuticals Inc.	-	0.002427184
	Milkhaus Laboratory, Inc.	-	0.002427184
	Sumitomo Chemical Co. Ltd.	-	0.001941747
	Boston Scientific Scimed, Inc.	-	0.001941747

This table lists the most frequent assignees within herbal and control patent groups. We see that firms that write both herbal and control patents are the most prolific set of firms. Given that the fraction of cap-exempt assignees in both the control and herbal patent groups are identical, and that a significant chunk of patents come from the set of assignees writing both herbal and control patents, we can infer that the patent matching procedure does well in matching assignees.

Robustness checks using context data

This section provides robustness checks for specifications (5)-(6) using an alternate measure contextual information.

Table A4. *Patent context and inventorship using most frequent herb*

	(1) Has Ethnic	(2) Has Ethnic	(3) Has Ethnic
$CONTEXT$ (Newest Herb)	0.0201^{***}	0.0162^{***}	0.0162^{***}
	(0.00584)	(0.00601)	(0.00563)
Constant	0.339^{***}	0.00344^{***}	-0.440^{***}
	(0.0224)	(0.00128)	(0.0312)
Time Fixed Effects	N	Y	Y
Controls	N	N	Y
Observations	2039	2039	2039
Adjusted R^2	0.011	0.022	0.124

Cluster robust standard errors in parentheses, at the assignee level. Dependent variable is an indicator for whether the patent contains Chinese/Indian inventors. Ethnic context measures the max inverse log frequencies of all herbs in a patent. There is a positive and significant association between the contextual

knowledge in herbs and the likelihood of having ethnic Chinese/Indian inventors on a patent, which is consistent with Chinese/Indian inventors writing patents with more contextual knowledge. The effect is robust for controlling for time fixed effects (Column 2 and 3), and the number of inventors and citations (Column 3).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5. *Patent context and inventorship during visa shock period*

	(1)	(2)	(3)
	Has Ethnic	Has Ethnic	Has Ethnic
$CONTEXT$ (Newest Herb)	0.0170***	0.00995	0.00998*
	(0.00592)	(0.00637)	(0.00575)
$TREAT$	0.0799***	0.416***	0.659***
	(0.0251)	(0.0452)	(0.0435)
$CONTEXT \times TREAT$	0.00765	0.0147*	0.0147**
	(0.00815)	(0.00812)	(0.00709)
Constant	0.307*** (0.0223)	0.00212 (0.00136)	-0.442*** (0.0311)
Time Fixed Effects	N	Y	Y
Controls	N	N	Y
Observations	2039	2039	2039
Adjusted R^2	0.016	0.023	0.125

Cluster robust standard errors in parentheses, at the assignee level. Dependent variable is an indicator for whether the patent contains Chinese/Indian inventors. Ethnic context measures the max inverse log frequencies of all herbs in a patent. There is a positive and significant association between contextual knowledge in herbs and the likelihood of having ethnic Chinese/Indian inventors on a patent, which is consistent with Chinese/Indian inventors writing patents with more contextual knowledge. The effect is robust for controlling for time fixed effects (Column 2 and 3), and the number of inventors and citations (Column 3).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Educational Background of Chinese/Indian Inventors

Table A6. *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 A7. *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 A8. *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

DDD additional tests and specifications

- **Estimates using inventor disambiguated data**

We present results using the inventor disambiguated data of Lai *et al*. (2013) in this section.

- **DD estimates using cap-exempt subset**

We can exploit further variation in the H1B visa cap-exempt employers to see its impact on herbal patenting behavior. We subset the data into two groups: patents with assignees that are exempt from the H1B cap, and those that are subject to the cap. While we expect similar coefficients for assignees that are subject to the cap, we do not expect to see results in the cap-exempt group.

Table A9 shows DD coefficient estimates for the cap-subject and the cap-exempt groups. Columns (1-2) use the cap-subject group and columns (3-4) are the cap-exempt group. Note that the number of cap-exempt assignees is significantly smaller than the number of cap-subject employees. While the coefficient on the DD estimate is positive and significant for cap-subject employers, it is statistically insignificant for the cap-exempt employers. Our estimates suggest that increasing the H1B visa cap raised the probability of herbal patents being invented by Ethnic inventors.

Table A9. *DD for cap-subject assignees vs. cap-exempt assignees, text match control*

	(1)	(2)	(3)	(4)
	Capped Assignee		Cap-exempt Assignee	
	Has Chinese/Indian	Has European	Has Chinese/Indian	Has European
HERB	0.0377	-0.130***	0.128*	-0.118**
	(0.0235)	(0.0211)	(0.0759)	(0.0577)
SHOCK	0.464***	0.719***	0.510***	-0.285***
	(0.0523)	(0.0398)	(0.130)	(0.106)
HERB x SHOCK	0.0885**	0.00516	-0.0820	0.0164
	(0.0420)	(0.0318)	(0.0983)	(0.0726)
Citations Count	Y	Y	Y	Y
Inventor Count	Y	Y	Y	Y
Constant	-0.443***	0.0923***	-0.395***	0.997***
	(0.0634)	(0.0354)	(0.0877)	(0.0676)
Observations	3219	3219	453	453
Adjusted R^2	0.121	0.030	0.079	0.065

Cluster robust standard errors at the Assignee level. We see significant effects on the interaction term only for capped assignees.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

- **Placebo test**

The main assumption to identification in a DD estimate is the common trends assumption: the control and treatment groups must have similar patterns in the dependent variable in the pre-treatment period. The literature has discussed how to test whether pre-trends align. As in Chetty et. al. (2009), we ran a permutation test to whether our common trends assumption holds. We randomly select a group of 2,060 patents to be our placebo herbal patents (treatment group), and also randomly select a consecutive 6-year period to be our placebo H1B visa shock, and we run the same specification as above, saving the coefficient on the DD estimate each time. We repeat this process for 200 different randomly selected groups of patents, and we plot the cumulative distribution function of the coefficients. Similar to a p-value, if the visa shock positively affected herbal patenting behavior, we would expect our coefficient to appear on the upper right tail of the cumulative distribution function.

Table A10. *Nonlinear specifications for triple differences model*

	(1)	(2)	(3)	(4)
Specification	OLS	Poisson	Logit	Conditional Logit (w/ Year FE)
Dependent variable: Has Chinese/Indian				
HERB	0.148*	0.496*	0.805*	0.797***
	(0.0814)	(0.270)	(0.433)	(0.257)
CAPPED	-0.0410	-0.236	-0.302	-0.301
	(0.0418)	(0.199)	(0.273)	(0.197)
TREAT	0.492***	15.76	12.95***	0 (.)
	(0.0753)	(13.13)	(1.084)	
HERB x TREAT	-0.117	-0.437	-0.642	-0.637
	(0.114)	(0.406)	(0.610)	(0.431)
CAPPED x TREAT	-0.0305	-0.192	-0.185	-0.186
	(0.0698)	(0.298)	(0.408)	(0.227)
HERB x CAPPED	-0.117	-0.281	-0.579	-0.573**
	(0.0833)	(0.283)	(0.448)	(0.290)
DDD	0.213*	0.746*	1.168*	1.161***
	(0.120)	(0.424)	(0.644)	(0.405)
Citations Count	0.000172	-0.0000166	0.00118	0.00114
	(0.000318)	(0.00188)	(0.00236)	(0.00234)
Inventor Count	0.0636***	0.140***	0.366***	0.363***
	(0.00638)	(0.0108)	(0.0431)	(0.0205)
Constant	-0.423***	-17.48*	-15.13***	
	(0.0614)	(9.918)	(1.069)	
Log likelihood	-2017.4	-2221.9	-1970.1	-1913.6

Estimation results for nonlinear models. Assignee level cluster robust standard errors in parentheses for (1)-(3), robust standard errors for (4). Note TREAT gets dropped for the conditional logit specification because within a year, there is no variation.

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

Herbal remedies - Relevance to western industry and scientific research

Table A11. *NBO Company Shares of Herbal/Traditional Products: % Value 2012-2016*

% retail value	**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