Centralization and Organization Reproduction: Ethnic Innovation in R&D Centers and Satellite Locations

William R. Kerr
Harvard University and NBER*

October 2023

Abstract

We study the relationship between firm centralization and organizational reproduction in satellite locations. For decentralized firms, the ethnic compositions of inventors in satellite locations mostly resemble their host cities, with little link to the inventor composition of their parent firms' R&D headquarters. For highly centralized firms, by contrast, organizational reproduction has an explanatory power equal to half or more of the host city effect. Reproduction is strongest when a firm exhibits a hands-on approach to the satellite facility, such as cross-facility team collaboration or internal talent mobility.

Keywords: Invention, patents, R&D, immigration, centralization, organizational reproduction.

JEL Classification: O32; F22, F23, J61, L22, L25, M51, O31, R11, R12

*I thank seminar participants, three referees, special issue editors, and Jasmina Chauvin, Ranjay Gulati, Toby Stuart, and Rodi Rodrigo Flores for helpful comments; Saheel Chodavadia and Ryan Barr for superb research assistance; and the National Science Foundation, Harvard Business School, the Smith Richardson Foundation, and the Ewing Marion Kauffman Foundation for financial support that made this research possible.

1 Introduction

Companies vary in the degree to which their R&D and innovation processes are geographically organized around a central R&D facility. Centralized companies group researchers together, promising enhanced information flow, greater opportunities for knowledge recombination across specialities, and less risk of knowledge leak to outside parties (Argyres and Silverman 2004). Geographically decentralized companies may provide firms better insights into multiple R&D communities, access to more diverse pools of talent, and aid in customizing technologies and products to local environments (Leiponen and Helfat 2011).

We study whether geographically centralized R&D systems exhibit greater organizational imprinting and reproduction (Stinchcombe 1965, 2000; Marquis and Tilcsik 2013) when staffing their satellite facilities compared to decentralized peers. While the design and centralization of firms has long been studied (e.g., Lawrence and Lorsch 1967; Allen 1977; Baldwin and Clark 2000), the relationship between centralization and organizational reproduction for peripheral facilities is unexplored. We compare how closely the ethnic compositions of inventors in satellite locations replicate the ethnic composition of inventors in their parent firms' R&D headquarters as well as the conditions and managerial actions that accentuate this reproduction.

Studying organizational reproduction through an ethnic inventor lens holds several advantages. Inventors typically work in teams (Wuchty et al. 2007) and often communicate or move across locations of a firm. This setting raises the prospect of organizational reproduction, and many studies document sorting, homophily, and discrimination on ethnic lines. Except for gender, which we also examine, other individual-level traits influencing affinity (e.g., religious affiliation, university attended) are usually weaker and/or unobservable. Additionally, significant differences in ethnic inventor populations have emerged across US cities since the 1970s. This heterogeneity aids in empirical identification and makes it possible to benchmark the magnitude of organization reproduction relative to local hiring.

Over the past 50 years, immigration has dramatically transformed the ethnic composition of US-based inventors. Most notably, inventors of Chinese and Indian ethnicity account for more than a fifth of US-based inventors today and are linked to the growing spatial concentration of patents in tech hubs like San Francisco and Boston.[1] Patenting firms must navigate these shifts

[1] San Francisco is the most dramatic example. Its share of US domestic patenting more than tripled from 5% in 1975-1984 to 18% in 2005-2014; contemporaneously, ethnic inventors grew to account for half of San Francisco's patenting. Consequently, in 1975, about one in every 220 US-based patents was invented or co-

in the geography of innovation. Many incumbents are rooted in cities distant from growing tech hubs, such as Procter & Gamble (P&G) in Cincinnati. Others, like Apple in San Francisco, have headquarters in a prominent tech hub but expand to other US cities. Firms rarely move their R&D headquarters but instead establish satellites with appropriate benefit-cost ratios (e.g., Shaver and Flyer 2000; Duranton and Combes 2006). Organizational reproduction may yield satellites for P&G and Apple that look more like their parent R&D headquarters than their host cities. We study this reproduction and its link to centralized R&D systems.

Our analysis considers 444 US public firms with significant patenting during 1975-2014. We identify the R&D centers of a firm ("RDHQ"). One location usually dominates, such as 3M in Minn.-St. Paul, Dell in Austin, Eli Lilly in Indianapolis, Ford in Detroit, Microsoft in Seattle, or Pitney Bowes in New York. For completeness, our procedure allows for up to three RDHQ facilities for organizations like IBM, General Electric, and Johnson & Johnson, but these cases are rare and not material. We measure R&D centralization as the firm's concentration of domestic inventive activity in the RDHQ. We quantify the ethnic compositions of inventors for satellite facilities, RDHQs, and their surrounding cities through inventor names.

There is a significant organizational reproduction on average. After controlling for local environments, a 10% increase in the share of ethnic inventors at RDHQ corresponds to a 1.5% increase at the satellite location. The explanatory power of this linkage to RDHQ is about a quarter of the explanatory power that host cities have. This effect is quite robust to sample design and estimation approach. Behind this average, organization reproduction is closely tied to R&D centralization. The satellites of decentralized systems mostly resemble host cities, with no evidence of organizational reproduction. By contrast, the most centralized firms show organizational reproduction that is comparable in explanatory power to half or more of the host cities' effect. The link between centralization and organization reproduction is stronger for satellites with easier travel from RDHQ and for satellites conducting similar technology work as RDHQ. We further quantify the roles of inventor collaboration and career mobility with RDHQ.

These specifications use between-firm variation for empirical identification, as R&D cen-

---

invented by an inventor of Chinese or Indian ethnicity living in San Francisco. In 2020, this share reached an astounding one in ten of US-based patents. This contribution of Chinese and Indian ethnic inventors in San Francisco was comparable to the collective patenting output of the 30 US states with the least patenting activity.

Estimates of the current share of patents and high-tech startups accounted for by immigrants in the US hover around 25% (Miguelez and Fink 2017; Ferrucci and Lissoni 2019; Kerr 2019; Bernstein et al. 2022).

tralization and RDHQ ethnicity are very persistent. One concern is that these traits may reflect other aspects of a firm and its strategy that also influence ethnic inventor compositions, potentially upwardly biasing our estimate of organizational reproduction. We test many potential issues directly (e.g., level of diversification, technology focus), but these extensions will always be incomplete. Thus, our final analysis embraces this persistence by using legacy workforces of firms to study future organizational imprinting, ideally removing the most worrisome issues. Isolating long-lived incumbent firms, the legacy ethnic composition of RDHQ inventors in 1975-1984 predicts well the ethnicity of inventors in RDHQ post 1995. Importantly, firm legacy also strongly predicts inventor composition in satellites after 1995.

This paper makes several important contributions. We extend the study of organizational reproduction to consider innovation satellites of firms and provide a novel technique to benchmark its importance. We also provide new insights into the relationship between firm centralization and organizational reproduction, directly testing for underlying mechanisms like collaboration and mobility. As levels of centralization are highly persistent for firms, we capture an important long-run organizational feature. We hope to spark further research into centralization's role for organization reproduction beyond innovation (e.g., production work).

Additionally, our paper contributes to immigration scholarship, although the organizational replication we describe extends more broadly (e.g., we show it holds with gender as well). Most studies analyzing immigration and invention treat the firm as a consolidated entity or use the firm's spatial layout to measure exposure to immigration shocks.[2] We contribute to a better understanding of immigration and the spatial response of firms. Prior work shows how an organization's employment of immigrants shapes overseas activity.[3] We also consider the firm's spatial R&D span, but our attention turns from the bridging power of ethnic networks to entering ethnically dissimilar places (e.g., Agrawal et al. 2008; Almeida et al. 2014). We also link to studies of how ethnicity and ethnic differences within firms influence innovation and firm performance (e.g., Hernandez and Kulchina 2020; Nguyen 2020). An important research topic going forward is how firms navigate uneven distributions of immigrants across cities.

[2] Examples include Kerr et al. (2015), Mayda et al. (2018), Bahar et al. (2020), Dimmock et al. (2021), and Doran et al. (2022).

[3] Examples from a wide literature include Rauch (2001), Saxenian et al. (2002), Singh (2005), Saxenian (2006), Docquier and Rapoport (2012), Foley and Kerr (2013), Hernandez (2014), Wang (2015), Breschi et al. (2017), Bahar and Rapoport (2018), Branstetter et al. (2018), and Glennon (2023). Balachandran and Hernandez (2021) consider venture investment decisions and performance.

2 R&D Centralization and Organizational Reproduction

We draw upon literatures that span organization studies, management scholarship, and economics. A first strand considers R&D centralization and its consequences. Argyres and Silverman (2004) hypothesize that centralized R&D systems decouple research efforts from the immediate commercial demands of divisions, reduce transaction costs, and clarify an organization's R&D objectives, consequently leading to higher-impact R&D that spans more technological domains. They argue decentralized systems instead better support product-specific innovation efforts, finding evidence that centralized systems in large companies produce more impactful patents as measured by citations. Lerner and Wulf (2007) subsequently link the highly cited patents in centralized R&D organizations to long-term incentive structures provided to R&D heads that have more organizational influence compared to peers in decentralized firms.

Leiponen and Helfat (2011) stress the use of multiple R&D facilities to acquire a broad base of external knowledge. Knowledge spillovers and absorptive capacity can be localized[4], and large patenting companies frequently seek an R&D presence in multiple cities. The authors theorize, and find evidence consistent with, multi-location R&D firms achieving more imitative rather than new-to-the-market innovation, consistent with Argyres and Silverman (2004).

Arora et al. (2014) study centralization from a control perspective, measuring the degree to which companies assign patents to affiliates versus corporate parents. The authors theorize that firms with centralized systems will invest more in internal research, relying less on external technology acquisition. When acquiring companies, moreover, acquisitions are more likely to be integrated. With some deviation from Argyres and Silverman (2004), the authors show control centralization links to internal orientation of activities across a large sample of patenting firms. Building a measure of R&D centralization through executive reporting structures in pharmaceutical firms, Eklund and Kapoor (2021) also identify how decentralized structures bring reduced internal resource flows and greater external sourcing.

In parallel, organizational scholars have long studied organizational imprinting and reproduction. Organizational imprinting (Stinchcombe 1965, 2000) details how organizations adopt and maintain traits of their initial environment and founders. Founding environments can strongly shape a firm's characteristics, and imprinted traits can persist despite subsequent

[4] Classics include Cohen and Levinthal (1990), Audretsch and Feldman (1996), Almeida and Kogut (1999), and Fleming et al. (2007). Nerkar and Paruchuri (2005) show inventor networks generate path dependency around central points in firms. Carlino and Kerr (2015) review the agglomeration and innovation literatures.

change in the surrounding environment. Past work identifies how organization imprinting influences the operating practices and strategic choices of firms.[5] Imprinting can be due to structural inertia within organizations (Hannan and Freeman 1977, 1984), limiting their plasticity. Organizational reproduction speaks to persistence of features, long-standing or newly acquired, within and across locations, such as demographics and inequality (Amis et al. 2020).

Related work documents homophily, or the tendency for people to form stronger connections with those they view as similar to themselves (McPherson et al. 2001; Lawrence and Shah 2020). Kanter (1977) finds managers are biased towards hiring candidates they see as like themselves. Research has identified homophily in multiple settings and types of decisions.[6] The consequences of homophily on performance are mixed (Ertug et al. 2022), in part because it can aid communication and efficiency but groups can suffer from lack of diversity.

Building upon the organizational imprinting/reproduction and homophily literatures, a first hypothesis anticipates organization reproduction in our ethnic inventor setting. While we are extending the theory to embody reproduction in satellite facilities, personnel in RDHQ frequently play a role in establishing a satellite and determining its staff, possibly even having RDHQ inventors re-locate to the new city. This influence is likely to be strongest when the new location is performing inventive work similar to the parent RDHQ.

Hypothesis 1: Satellites reproduce the ethnic inventor composition of their parent RDHQ.

Despite this research progress, how centralization influences organizational imprinting and reproduction is less explored. Waeger and Weber (2019) consider how centralization affects organizations' reaction to changing environments that conflict with existing organizational logics. In a gender context, Kleinbaum et al. (2013) measure how homophily is greater within organizations than across them and depends upon firm size.

The degree of control embedded in centralized R&D systems (Eklund 2022) raises the likelihood of organizational reproduction. This reproduction will likely be evident in more ethnically homophilous workers at the satellite location given the higher influence of RDHQ on inventor staffing, processes, and operations at the new satellite in centralized R&D systems. The increased level of organization reproduction with centralization is especially likely when

[5]Examples include Kimberly (1979), Schein (1983), Boeker (1989), Bamford et al. (2000), Baron et al. (2001), Kriauciunas and Kale (2006), Burton and Beckman (2007), and Marquis and Tilcsik (2013).

[6]Examples include Byrne (1971), Williams et al. (2012), Hegde and Tumlinson (2014), Rivera (2015), and Ruiz Castro and Holvino (2016).

satellites and RDHQ are engaged in technically similar work given the greater attention that RDHQ may devote to the peripheral facility.

Hypothesis 2: Centralized R&D systems develop satellites that more closely reproduce the ethnic inventor composition of the parent RDHQ than decentralized peers.

A few studies explore how internal linkages shape activity across R&D locations. Firms with strong internal linkages and intrafirm mobility share knowledge efficiently across facilities (Lahiri 2010). Alcacer and Zhao (2012) further show that strong internal linkages reduce the threat of knowledge expropriation in multi-location structures by allowing stronger effective control and shifting innovations towards activity that holds greater internal value than outside. They find more cross-cluster inventive teams when facilities are located in clusters with many competitors. Choudhury (2017) provides an important additional mechanism by recognizing the role of personal contact for resource access. Using internal records of a multinational and a clever experiment, he shows that intrafirm mobility brings satellite inventors face-to-face with individuals who allocate resources, boosting innovation outcomes.

Studies of talent mobility inside organizations consistently find mobility aids knowledge sharing and resource access, while also reinforcing organizational patterns and practices. The same holds for cross-location team collaborations. As Choudhury (2017) further articulates, the importance of mobility holds in a knowledge-based view of the firm, where information resides in individuals to capture, or in an information processing view of the firm.[7]

Engagement with satellites extends RDHQ's influence. As later documented, centralized systems display higher intra-firm mobility and use of spatially distributed inventive teams. These interactions can promote organization reproduction through the direct mobility of individuals from RDHQ to the new location. Engagement with RDHQ can also promote homophily among new hires at satellites. We anticipate that centralization will mostly link to organizational replication when these intra-firm activities are present.

Hypothesis 3: Intra-firm mobility of inventors and spatially distributed inventive teams are important mechanisms linking centralized firms to organizational reproduction in satellites.

[7] See Madsen et al. (2003), Singh (2008), Karim and Williams (2012), Choudhury (2016, 2022), Kerr and Kerr (2018), Choudhury and Kim (2019), Marino et al. (2020), Choudhury et al. (2022), and Castellani et al. (2022).

3 Data and Methods

3.1 Data Collection and Preparation

We downloaded all patents granted by the United States Patent and Trademark Office (USPTO) during 1976-2020 through PatentsView. Our analysis groups patents into four time periods by application years: 1975-1984, 1985-1994, 1995-2004, and 2005-2014. The appendix discusses the use of application years and our development of consistent technology codes. Most of our estimations employ the NBER system (Hall et al. 2001), which has 6 main technology categories and 36 subcategories. Patents list the cities of inventors, enabling discernment of satellite locations. We group cities into 281 Metropolitan Statistical Areas (MSA) with constant boundaries. We use consolidated MSAs such that San Francisco includes San Jose, Oakland, etc.

Li et al. (2014) pioneered techniques to link inventor names on patents into longitudinal inventor identifiers, correcting for minor spelling differences or nicknames. PatentsView provides inventor IDs based upon the USPTO's continuation of this effort. We utilize these inventor IDs with modest modifications noted in the appendix. We identify inventor ethnicity using a name procedure first developed by Kerr (2007) and Kerr and Lincoln (2010). The procedure uses multiple databases of common ethnic names, initially assembled for marketing purposes, and name conventions by ethnicity. For example, the surnames Patel and Gupta link to the Indian ethnicity, and Rodriguez and Hernandez map to the Hispanic ethnicity. The procedures provide ethnicities for more than 99% of US-based inventors. Identified ethnicities include Anglo-Saxon, Chinese, European, Hispanic, Indian/South Asian, Japanese, Korean, Russian, and Vietnamese.

Figure 1 shows the share of domestic patents accounted for by US-based ethnic inventors over time. Anglo-Saxon and European ethnicity inventors, who are not shown on this figure, decline from a combined 90.6% of invention in 1975 to 66.0% for 2019. Ethnic inventors, defined as those not of Anglo-Saxon and European heritage, grow to one-third of US patenting; Chinese and Indian ethnic invention surge from collectively 3.4% of 1975 patenting to 22.3% for 2019. Appendix Table 1a provides further descriptive statistics.

Finally, we group patent assignees into companies. The USPTO assigns patents to entities as listed on applications, which do not map immediately into a company. Issues include misspellings, acronyms (e.g., "IBM"), and subsidiary relationships. PatentsView's algorithms correct basic challenges, and we further adopt the Discern Data (Arora et al. 2021a,b) that aligns patent assignees into Compustat-linked, public parent companies. Our primary sample with this "ultimate owner" firm represents the most comprehensive corporate span possible.

3.2 Panel Development

Our panel balances including as many firms and locations as possible with avoiding frivolous cases (e.g., a remote inventor collaborating once with a company). We use the term "firm" to designate the "ultimate owner" firm in Discern. "Inventor-patents" signifies a US-based inventor paired with a specific patent. Thus, if a patent has four inventors based in the US and one inventor abroad, it forms four inventor-patents in our data frame.

The first step requires a firm have at least 250 patents during 1975-2014 that include US-based inventors living in an MSA. We also require that a majority of the firms' patents are of domestic origin, thus excluding foreign firms conducting most of their patenting abroad. These steps retain over 90% of the patents linked to public companies. We finally require that included periods for a firm have at least 100 inventor-patents.

We next identify RDHQs. Our baseline approach identifies for each period the top MSAs for the firm such that they account for 30% or more the firm's US-based inventor-patents. We collect these MSAs across periods for a firm to define up to three RDHQs. This approach ensures consistency over the sample period and the inclusion of very decentralized firms.[8] Most firms have a single dominant RDHQ that never changes, and our results are robust to considering unambiguous cases where a single RDHQ always holds more than half of the firm's patents. Our primary measure of centralization, labeled $RDCentr_f$, is the share share of a firm's domestic patents in its RDHQ in the first period observed. This metric is similar to studies that use shares of R&D budget or technical staff in a lead city (e.g., Png 2019). We prefer an initial, time-invariant measure of R&D centralization, but we observe very similar results when using average, contemporaneous or lagged centralization due to its persistence.[9]

We define satellites at the MSA level, which in some cases aggregates establishments. We have 30,564 unique firm-MSA-period observations that exclude RDHQs. Most of these are

[8] For most firms, a single MSA is always the dominant RDHQ. In some cases, the main RDHQ location switches once, such as Intel and the establishment of its Portland R&D campus. Finally, some firms like IBM and Johnson & Johnson are sufficiently decentralized such that multiple MSAs are needed to reach the 30% threshold. We drop 12 firms where 4+ MSAs would have been required to reach 30%.

[9] A regression of R&D centralization on firm and technology-period fixed effects has an R-Squared value of 0.91. For RDHQ ethnic worker shares, the equivalent R-Squared is 0.93.

small and ephemeral, perhaps capturing entering/departing workers or one-off collaborations. We thus require locations have at least five inventor-patents during a period, forming a final sample of 12,222 unique firm-MSA-periods that account for 30.6% of Discern patents.

Table 1 reports that the average satellite location has 67.8 inventor-patents. The average ethnic inventor share is 16.9%, typically falling below expectations based upon the host city. 29% of satellites are more than 1500 miles away from their RDHQ, and 54% focus on a different technology. Locations show important collaboration and intra-firm mobility with RDHQ. The RDHQ locations have an average ethnic inventor share of 23.6%, slightly above host cities. RDHQs maintain an average of 55.8% of the firm's overall US-based patents; initial centralization is 64.7% on average, with a range of 31.6%-98.3%.

While using data in multiple ways could introduce a common method bias, this project is secure as ethnic inventor composition and centralization features are orthogonal to each other. Appendix Table 1b provides a correlation matrix to this end. Independent data are also used for measuring the ethnic compositions of each facility and location. We hope future research can assemble other types of R&D data for satellites, but this is difficult as surveys of R&D activity are almost always firm-level in scope.

3.3 Estimation Approach and Specification

Our baseline regressions quantify the degree to which the ethnicity of the inventive workforce at a satellite reflects the contemporaneous ethnicity of the firm's RDHQ. For a satellite i that is located in MSA m during period t, we consider the specification

$$\begin{aligned} ETH_{i,m,t} &= \beta ETH_{f,t}^{RDHQ} + \gamma_1 ETH_{i,m,t}^{Exp} + \gamma_2 ETH_{f,t}^{Exp-RDHQ} \\ &\quad + \eta_m + \eta_{c_f,t} + \gamma P_{i,m,t} + \delta P_{f,t} + \epsilon_{i,m,t}, \end{aligned} \tag{1}$$

where firm f signifies the parent firm.

The outcome variable, $ETH_{i,m,t}$, is the observed ethnicity of the inventive workforce at the satellite aggregating over inventor-patents. It is the average share of the satellite's inventive workforce of neither Anglo-Saxon nor European ethnicity. The variable ranges continuously from 0% (all inventors are of Anglo-Saxon/European ethnic descent) to 100% (all inventors are of other ethnic descent). Our focus is on the regressor $ETH_{f,t}^{RDHQ}$ that measures the contemporaneous ethnicity of the inventive workforce in firm's RDHQ. It is measured in the same way as $ETH_{i,m,t}$ but using inventors located at RDHQ.

To isolate organizational reproduction, our specification must first account for the tendency of ethnicities to concentrate in particular technologies. For example, Appendix Table 1a shows significant over-representation of Indian ethnic inventors in computers. Without appropriately accounting for this clustering, we risk over-estimating organizational reproduction if the RDHQ and satellite are engaged in similar work. To address, we build detailed control variables of expected ethnic shares. Specifically, $ETH^{Exp}_{i,m,t}$ is the expected ethnicity of a satellite based upon invention being conducted in the surrounding MSA during the same time period. We calculate this measure by replacing the observed ethnicity of each inventor-patent with the contemporaneous ethnic average of the host city in the same NBER tech subcategory. We include the satellite location itself in these local averages. We similarly control for the expected ethnicity of the parent firm RDHQ with $ETH^{Exp-RDHQ}_{f,t}$, following the same construction. These controls flexibly capture any deviations in ethnic representation by city and technology.

We also include a vector fixed effects η_m for the MSA of the satellite location. As our expected ethnicity measure $ETH^{Exp}_{i,m,t}$ is constructed specific to a satellite location's activity in a period, we can identify its role even in the presence of the MSA fixed effects. We further include a vector of fixed effects $\eta_{c_f,t}$ that interact the primary NBER technology category of the firm with each time period to remove macro dynamics. We finally control for the log patent count of the satellite location ($P_{i,m,t}$) and parent firm ($P_{f,t}$) in the period. Estimations cluster standard errors by parent firm and are unweighted.

Beyond accounting for ethnic differences across technologies, endogeneity concerns exist. A firm seeking to enter a new city with a different ethnic inventor profile than its RDHQ may hire new workers at RDHQ that resemble the target location. Additionally, staff sourced from satellites may be redeployed to RDHQ. These forms of reverse causality would upwardly bias our estimates of organizational reproduction. Additionally, omitted factors like a firm-wide effort to promote hiring more ethnic talent could exist. Specification (1) does not address these concerns, and we later consider legacy workforces for long-lived firms to make progress.

We utilize two approaches when considering R&D centralization and its link to organizational replication. A graphical approach replicates specification (1) using 55 sub-samples of our data organized by the initial R&D centralization of the firm. The appendix provides further details. A second approach is parametric, specifying $RDCentr_f$ has a linear effect. In these estimations, we extend specification (1) to control for $RDCentr_f$ and have interactions of $RDCentr_f$ with $ETH_{f,t}^{RDHQ}$ and $ETH_{i,m,t}^{Exp}$. We discuss the interpretation of R&D centralization further below.

4 Results

4.1 Full Sample Analysis

Table 2 investigates Hypothesis 1 using the full sample of satellites. The first column shows a baseline result with specification (1). The β coefficient captures a strong link of ethnic invention at the satellite to the ethnic invention at the firm's RDHQ. A 10% increase in the share of ethnic inventors at RDHQ corresponds to a 1.5% increase in ethnic invention at the periphery location. The interquartile range of the ethnic share at RDHQ is 17.5%, consistent with a 2.6% higher ethnic share at the satellite, relative to a sample mean of 16.9%.

Not surprisingly, the γ_1 coefficient also shows a tight connection of the focal satellite's ethnic workforce to the inventive community conducting similar patenting in the host city. A 10% increase in the expected ethnicity of the satellite location is correlated with a 6.1% increase in observed ethnic invention. Thus, the RDHQ workforce has about a quarter of the explanatory power that is observed for the local inventor population (0.149/0.612 = 24.3%), which is remarkable in size. The interquartile range of the expected ethnic share of a location is 16.0%, consistent with a 9.8% increase in observed ethnic invention. By contrast, the γ_2 coefficient shows no additional link of expected ethnic contributions in the city of RDHQ.

Appendix Tables 2a-2d report variations. The results are remarkably robust, and we note here just a few features. First, our preferred specification estimates that RDHQ can account for an effect on satellite workforce equal to 24% of the satellite's host city effect. Depending upon the stringency of fixed effects deployed, explanatory power ranges from 19% to 27%. Across all of the specifications in the Appendix, the relative effect averages 25.3% with a standard deviation of 5.7%. Reproduction is comparable when splitting the sample by patent filing decentralization or the level of firm diversification. Finally, the results are similar when splitting the sample based upon technology growth rates and ethnic inventor growth rates.

The remaining columns in Table 2 consider sample heterogeneity. Columns 2 and 3 demonstrate that organizational reproduction is higher for satellites that are within 500 miles of RDHQ or have a direct, commercial plane flight available.[10] Appendix Table 2e displays more

[10] We appreciate Bernstein et al. (2016) providing data on commercial flights.

granular distance bands. The most significant effects are evident when satellite locations are within 100 miles of RDHQ, with a β coefficient of 0.337 $(0.095)^{+++}$ that is more than double the strength of the sample average. This link to RDHQ decays to half its initial magnitude when the satellite location is 500 or more miles away. At all distance bands, decay is also faster without a direct flight to RDHQ.

Columns 4 and 5 show stronger effects when satellites have the same technology focus as RDHQ, defined using modal NBER subcategories of inventor-patents for the two locations. The differences are consistent with our theoretical hypotheses and are also evident at more detailed technology levels.[11] By contrast, we do not observe large differences in organizational reproduction when splitting the sample using broad NBER categories. The match of technologies across locations within the firm is the key feature linked to organization reproduction, not technology domain.

Appendix Table 3 explores additional demographic variation. We first segment Chinese and Indian ethnic inventors from other ethnic groups, documenting deeper levels of organization reproduction that are ethnic-group specific. The link for Chinese and Indian ethnic inventors appears stronger than for other ethnic groups, reflecting their primary role in the growth evident in Figure 1. While the rest of our study continues with the broader measure of non-Anglo-Saxon and non-European ethnic inventors, these analyses suggest organization reproduction is happening on more finely grained dimensions as well. The appendix also shows that a gender dimension to organizational reproduction coexists with the ethnic dimension, perhaps even being somewhat stronger. A higher share of female inventors in RDHQ corresponds to greater female representation among inventors in satellites. This extension serves as a robustness check on our ethnic focus, but it also further highlights the general nature of organization reproduction.

4.2 R&D Centralization and Organizational Reproduction

We next consider Hypothesis 2 regarding R&D centralization and the level of organizational reproduction for satellites. Figure 2 replicates specification (1) using sub-samples of our data organized by the initial R&D centralization of the firm. These estimations show non-

[11] 3379 satellites have the same primary USPC technology as RDHQ, compared to 5578 at the subcategory level. We obtain β coefficients of 0.219 $(0.054)^{+++}$ and 0.114 $(0.032)^{+++}$ when splitting the sample for same versus different USPC technology, respectively. Thus, a tighter technology match increases the link observed.

parametrically the explanatory power of RDHQ's inventive workforce for the inventor composition of satellites, and we continue to compare estimates to the satellite's host city.

For decentralized firms with low values of $RDCentr_f$ near to the 30% threshold, the ethnic compositions of RDHQ and satellites show very limited connection, consistent with minimal organizational reproduction. At higher levels of concentration, however, the link is quite strong. Columns 6-8 of Table 2 disaggregate the sample by $RDCentr_f$: (30%,60%], (60%,75%], versus (75%,100%). Among the more decentralized firms, a β coefficient of 0.066 $(0.040)^{+}$ emerges. Indeed, the β coefficient across the (30%,45%] range is -0.012 (0.062), compared to 0.088 $(0.046)^{+}$ in the (45%,60%] range. For higher levels of initial centralization, the relationship is quite strong and precisely estimated. The β coefficient is 0.242 $(0.056)^{+++}$ for the (75%,100%) range. While the link to RDHQ remains smaller than the host city's influence, the convergence for centralized firms is remarkable.

Complementing this visual analysis, the first column of Table 3 extends our baseline specification to include interactions of our main regressors with $RDCentr_f$. The main effect for the Ethnic share of RDHQ confirms again that decentralized firms display minimal organizational reproduction. For each 10% increase in initial R&D centralization, we measure a 0.033 stronger tie. The bottom of Table 3 provides linear combinations of the main effect and interaction term at 30%, 60%, and 90% centralization. For the most centralization firms, a 10% increase in ethnic invention at RDHQ corresponds to a 2.2% increase at satellites.

We earlier split the sample by distance and technology focus to test hypotheses. Figure 3 repeats our graphical analysis with these splits, and Appendix Table 5a contains the corresponding interaction estimations. The evidence for strong organizational reproduction grows with R&D centralization for both proximate and distant satellites. For proximate locations, the explanatory power of the RDHQ equals that of the host city when centralization is above 80%. For distant locations, the gap does not close. While visually the relationship of organization reproduction to R&D centralization appears stronger for proximate locations, we cannot reject the null hypothesis that these series grow at the same rate across the full distribution of R&D centralization.

The technology split is similar. Except for extremely low values of R&D centralization, we consistently observe evidence of organizational reproduction for satellites operating in the same technology class as RDHQ. At high values of centralization, RDHQ has comparable explanatory power to that of the host city when the satellite is in the same technology area. By contrast, satellites with a different technology focus show limited evidence of organization reproduction. In this split, we can reject the null hypothesis that the growth of organization replication with R&D centralization is the same.

Appendix Table 5b shows these results are robust to using lagged measures of RDHQ centralization. We find quite similar outcomes when using initial centralization of the firm, lagged centralization of the firm, or the average centralization over periods of the firm. In all variants, the evidence for organizational reproduction is greater in more centralized firms. This stability is due to the extreme persistence of centralization and our use of between-firm variation.

While these results are quite sharp, two interpretation questions regarding centralization exist. One issue returns the omitted variables noted in Section 3.3. Our measure of R&D centralization may be correlated with other strategic and operational choices of the firm beyond size and technology focus that possibly influence the ethnic similarity over locations. Our best approach on this front remains a battery of robustness checks, and Appendix Tables 6a-6d repeat our interaction analysis akin to Appendix Tables 2a-2d. Encouragingly, we find quite comparable results across specification checks, including sample splits by filing decentralization and firm technology diversification.[12] The results are even stronger when using Coarsened Exact Matching to pair firms above and below median centralization by RDHQ ethnicity, inventor-patent counts of satellite, inventor-patent counts of parent firm, and technology-period. However, these tests are only a partial solution.

Additionally, while this metric of centralization is commonly used, most studies do not explore why the link exists. Many mechanisms could underlie its importance: for example, talent mobility, formal approval processes, hierarchical cultural norms, and dense within-firm networks. The next subsection analyzes two mechanisms traceable in our data, realizing we do not observe others that could also play a role. This raises the importance of complementary single-firm studies like Choudhury (2017) to assess firm policies and similar.

[12] The most notable difference in Appendix Tables 6a-6d is the assignee-level estimation. While the link of satellites to RDHQ is evident at 30%, 60%, and 90% centralization levels, the linear term is significantly weaker.

4.3 Inventor Collaboration and Mobility

Research consistently demonstrates that team collaborations and mobility of inventors are important for productivity and career development (e.g., Akcigit et al. 2018; Miguelez 2019; Moretti 2021; Tzabbar et al. 2021). Hypothesis 3 proposes that inventor collaboration and mobility are also important mechanisms for organizational reproduction. We next document evidence regarding these mechanisms and their connections to R&D centralization.

Centralized R&D systems typically display greater linkages across their (typically fewer) inventive locations than decentralized systems. Tables 4a and 4b document some of these magnitudes. A 10% increase in $RDCentr_f$ correlates with 3.8% increase in team collaboration on patents between the satellite and RDHQ, relative to a baseline rate of 30.1% collaboration. The comparable effect of a 10% increase in $RDCentr_f$ for inventor mobility between the satellite and RDHQ is 1.9%, versus a sample average of 10.1%.[13] Additionally, patents developed in the satellites show a 2.7% higher rate for citing prior work in RDHQ for each 10% growth in R&D centralization.

Recognizing these stronger connections over facilities for centralized systems, Table 3 and Figure 4 explore whether organizational reproduction is only evident when these interactions across parts of the firm are also evident. We first split the sample by period into halves based upon median rates of observed inventor collaboration between satellites and their parent RDHQs. Collaboration appears quite important and also linked to the larger influence of RDHQs evident in centralized systems. On the left hand side of Figure 4, there is limited evidence for organization replication among decentralized systems, independent of collaboration rates. The strongest links exist among centralized systems with above-median levels of collaboration. Table 3 shows that the dependency of organizational reproduction on higher levels of centralization is statistically different between the two samples. When looking at inventor mobility, we observe similar linkages.

In Tables 5a and 5b, we take an important additional step in showing that this organization replication extends to inventors who are not personally collaborating with RDHQ nor moving over facilities. This carryover effect is consistent with the links of RDHQ and the satellite

[13] Mobility is defined for inventors based upon whether they patent while working at RDHQ at least once during their career. A career-level definition is simpler in the presence of out-and-back movements from RDHQ displayed by some inventors (even within a decade period). Mobility connects more experienced inventors to the satellite.

being sufficient to transfer practices that shape future organizational reproduction, but there may be other unobserved links too. In auxiliary estimations, we further consider whether imbalances in patenting experience across inventors shape the interactions, such that having a more experienced inventor on the collaborative team in the RDHQ leads to stronger effects. We do not observe important differences in this regard.

4.4 Legacy Sample Analysis and Imprinting

The results in the full sample analysis speak to a tight linkage of the ethnicities of inventors at a firm's RDHQ and satellites. Beyond the controls built into model (1), the multiple robustness checks and extensions suggest the link is unlikely to be spurious. Yet, the contemporaneous nature of the exercise and use of between-firm differences significantly limit claims of causality. Moreover, firm R&D centralization is not random, and choices about centralization may have been driven by unobservable firm strategic choices.

An exploration of the history of organizations helps close this gap, providing insight into early organizational imprinting. We examine how early choices about staffing and centralization persist over subsequent decades and connect to future organization reproduction decades later. To the degree that initial factors influencing strategic choices fade with time, this technique removes the most worrisome endogeneity. We test whether legacy ethnicity of inventors at RDHQ during 1975-1984 predict the future ethnicities of inventors at satellites. Our interest in RDHQs is due in large part to their central position in the firm and their long histories compared to the satellite locations. The full sample includes companies like LinkedIn and Salesforce.com that were founded after 2000, thus having limited corporate history, and others from the 1970s and 1980s that went out of business. This legacy sample is selected on survival but allows a longer time frame for the study of firms and imprinting.

Our legacy sample requires companies be active in patenting during the 1975-1984 initial period and have satellites in the two decades after 1995, all the while maintaining a lead RDHQ. The inclusion requirements remain otherwise the same as described for the full sample. This cull leaves 4142 observations from 126 firms. These observations account for 62.7% of the patenting that was captured by the full sample for 1995-2014. Appendix Table 7 provides descriptive statistics on this sample, which are mostly comparable to the full sample despite a lower share of ICT firms.

Table 6 commences by measuring the persistence of initial inventive workforces at RDHQ itself. Panels A and B look at 126 and 83 legacy RDHQs present in 1995-2004 and 2005-2014, respectively. Column 1 regresses the total ethnic composition of the inventive workforce in the later periods on the legacy ethnic share in 1975-1984. Regressions are unweighted, include technology fixed effects, and report robust standard errors. The relationship is quite strong, measuring that a 10% higher ethnic workforce in 1975-1984 at RDHQ links to a 7.3% and 6.3% higher ethnic inventor share in the two periods after 1995, respectively.

Columns 2-5 then split the workforce into inventors who first patent with the firm in 1975-1984, followed by each subsequent decade to 2005-2014. Column 2 is comprised of inventors who were present when the legacy workforce was estimated, but it is not colinear with the legacy regressor in that some inventors have left the firm and others changed their rate of patenting. Individuals in Columns 3-5 are by definition not part of the legacy ethnic measure, as they first patent with the firm after 1984. In some cases, they may have been employed at the firm in the initial years, and only started patenting later, but that becomes increasingly unlikely with later start dates. In 1995-2004, RDHQs average about 10% of their inventive workforce being individuals who first patented in 1975-1984; this average further declines to under 5% for invention being done in 2005-2014.

We observe significant persistence to later cohorts, with a 10% higher ethnic workforce in 1975-1984 at RDHQ predicting a 3.7% higher ethnic inventor share among individuals who first patent with the organization in 2005-2014. There is decay over time within each panel, but change is slow. This documents why there is insufficient variation for a panel analysis of RDHQ ethnicities of RDHQ, and so we instead consider how well the past can predict the ethnic composition of future satellite locations.

Table 7 documents the persistence estimations akin to Table 2. We model the ethnic share of RDHQ during 1975-1984 as the explanatory variable, with all other features following specification (1). There is a considerable link of the ethnic legacy of RDHQ to the inventor workforces in satellites in future periods, providing additional support for Hypothesis 1. A 10% increase in the ethnic share of the RDHQ's workforce in 1975-1984 predicts a 2% increase in the ethnic share of satellites after 1995. This explanatory effect is equivalent to 30% of the host city effect. In contrast to the full sample, however, Columns 2-5 show limited differences in sample splits by distance or technology focus. Columns 6-8 show evidence for organizational reproduction again rises in the level of initial R&D centralization, supporting Hypothesis 2. Appendix Figure 1 provides our non-parametric analysis of legacy workforce and levels of initial R&D centralization.[14]

One can contemplate running the estimation as an IV framework. Column 1 of Table 6 would be akin to the first-stage regression, and Table 7 would be the reduced-form results. When doing so, the IV estimate would be 0.297 (0.086)$^{+++}$ with an OLS of 0.149 (0.053)$^{+++}$. While comforting to report, we believe it best to model the legacy analysis through reduced-form persistence as it remains possible that other factors correlated with 1975-1984 ethnic invention (above and beyond our controls) are influencing the future satellite facilities. Appendix Figure 2 shows the comparability of OLS and reduced-form explanatory power.

We have also confirmed Table 7's results in several other ways. The legacy effect is stronger for more recent and smaller satellites opened by the firm. These results are consistent with a more substantive imprint from RDHQ and its legacy on the launch of satellites, weakening over time with operation and local growth. Similar to the full sample, the legacy effect is sharper when the RDHQ is located outside of the major US tech clusters (Kerr and Nicoud 2020). Finally, the role of the legacy inventor ethnicity persists when restricting the sample to firms for which workers first patenting in 1975-1984 make up less than 10% of the contemporary inventor pool at RDHQ.

5 Discussion and Future Research

Since 1975, immigration has more than tripled the share of US-based patenting being conducted by ethnic inventors. This rising tide differentially affected cities (e.g., Peri et al. 2015) and thus incumbent firms, providing a laboratory to examine organization reproduction. In contrast to studies emphasizing the advantages of ethnic affinity for the placement and success of overseas R&D centers, we consider the development of domestic satellite locations that are spatially separated from the dominant R&D location for a US-based company. Many incumbents have rich legacies and cultures that may shape their satellites whatever the host city, and we explore how the resulting satellite reflects the influences.

Consistent with organizational replication, the ethnicity of inventors at a firm's RDHQ

[14] Splitting the legacy sample becomes strained by the smaller number of observations and many fixed effects. Appendix Table 8 shows very similar results to Table 7 when relaxing technology category-period fixed effects to be period fixed effects.

predicts well the ethnicity of inventors working at satellites. Our preferred estimation suggests this link has 24% of the explanatory power that hiring from the host city's inventor pool holds. Evidence for organizational reproduction is remarkably robust across specification variants, intuitively growing with more proximate sites and when satellites are inventing in the same technologies as RDHQ. Given that ethnicity is a salient trait (often to the point of discrimination) and inventors usually work in collaborative teams, we suspect that this effect would be at the higher end of organizational reproduction along dimensions of affinity.

Evidence for organizational reproduction is particularly strong for centralized R&D systems, especially above initial concentration levels of 60%. This is an important connection of the two concepts for the literature. Decentralized systems display minimal to no organizational reproduction, mostly looking like their host cities. Evidence for organizational reproduction is much stronger at 60% centralization, and the implied magnitude of RDHQ's influence almost further doubles when at 90% centralization.

Beyond this measurement, a strength of patent data is that we can observe mobility and teamwork. Inventor mobility within the firm between the satellite and RDHQ provides the strongest tie, closely followed by inventor collaboration spanning locations. Direct interactions (i.e., the movers or the collaborators) boost the ethnic similarity of the satellite to RDHQ the most, but the ethnic similarity of others at the satellite are also closer to RDHQ when these interactions are present. This is consistent with unobserved features like culture and processes also being shaped by these interactions with RDHQ.

Centralized firms display more inventor mobility and collaboration than decentralized peers. Importantly, absent these inventor-level linkages, we do not observe significant organization reproduction among centralized firms. This difference helps verify our approach and characterization of centralization. It also highlights channels that other studies of R&D centralization and its organizational impacts may want to consider.

There are several important limitations and areas for future work. These results are context dependent, focused on established and relatively successful US-based firms. Many of these firms would have aspired for ethnic diversity and access to more knowledge inputs into the firms. This is particularly true for our legacy sample analysis, as the firms have survived several decades. Future research needs to explore settings with different conditions. Features of organizational reproduction beyond the workforce composition (e.g., how hierarchical a satellite location is) are also worthy of study and may differ from the demographic composition we study here. Contrasting alternative forms of R&D centralization (Eklund 2022) may also be possible.

Second, we fall short of establishing a causal relationship. The phenomenon is slow-moving, and organizations like P&G, General Motors, and Oracle have persistent corporate identities. Our analyses accordingly use between-firm variation. We show the explanatory power of inventive workforces at the R&D headquarters during the mid 1970s and early 1980s on the composition of satellite workforces decades later. The persistence and reproduction of these initial conditions are consistent with a causal story, but we cannot rule out additional unobservable features that could influence the findings.

An open question is the consequences of these phenomena for organizations. The appendix provides some initial exploration using patent traits. When comparing satellite locations of a firm, satellites with links to RDHQ show higher patent novelty, stock market response to patent announcements, and internal citations. These results are preliminary because we have not identified how to model realized organizational reproduction, instead showing that the mechanisms linked to organizational reproduction are also linked to better patents. To take the next step, research needs to devise a method to consolidate the dimensions that we have been comparing to each other at the aggregate level (i.e., the measured link to RDHQ versus link to host city) into a single metric that can be applied at the satellite level.

Future work can explore more interactions across locations. Patent citations and component patent sub-classes could be useful for exploring how knowledge is moving between the R&D sites of the firm and the conditions that shape whether a satellite is dependent on RDHQ or independent. Our initial results on technological distance between locations and the mobility of inventors suggest there is significant heterogeneity, possibly even across locations within a firm. Work to bring additional structure to these important design considerations would be valuable, as would their interactions with the quality and career development of inventors.

Similarly, while best conducted for specific industries, future work can combine data on firm strategies with the study of organization reproduction. Patent data do not reveal why any specific facility was launched and best measure average treatment effects. A great opportunity is to combine heterogeneity in firm strategy with analysis of which cities are chosen for facilities and how they are staffed and managed. Indeed, this type of work already exists when analyzing overseas entries by US firms, and we believe the growing heterogeneity across cities within countries warrants attention, too.

We also hope to consider inorganic acquisitions by firms of R&D centers in other locations. Discern's parent firms provide broad insight on corporate structures, and results are comparable using assignees only. But more can be done to test the implications of M&A activity. Walmart entered San Francisco in 2011 to form @WalmartLabs through the purchase of Kosmix, a business founded by two Indian entrepreneurs. The acquisition provided Walmart a strong physical presence and early talent. This type of inorganic entry may provide a more distinctive ethnic composition of the new facility compared to RDHQ, perhaps a tactic even purposefully chosen to make the organizational bridge. The strategic and organizational factors that shape the choice to send inventors from RDHQ to seed a satellite location, hire directly new talent there, or acquire a local business like Kosmix are important to study.

Finally, this paper focuses on invention, and it is important for future research to evaluate the aggregate firm implications of these phenomena, including impact on physical operations and competitiveness. A frontier of studies jointly models firm location decisions with respect to knowledge development and production facilities, such as Tecu (2013), Alcacer and Delgado (2016), and Delgado (2020). Likewise, employer-employee data may allow a contrast of ethnic heterogeneity in operational and front-line workers of a firm with the inventors across sites.

References

Agrawal, A., D. Kapur and J. McHale. (2008). "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data." Journal of Urban Economics 64(2), 258-269.

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva and V. Sterzi. (2018). "Dancing with the Stars: Innovation Through Interactions." NBER Working Paper 24466.

Alcacer, J. and M. Delgado. (2016). "Spatial Organization of Firms and Location Choices through the Value Chain." Management Science 62(11), 3213-3234.

Alcacer, J. and M. Zhao. (2012). "Local R&D Strategies and Multilocation Firms: The Role of Internal Linkages." Management Science 58(4), 734-753.

Allen, T. (1977). Managing the Flow of Technology. MIT Press.

Almeida, P. and B. Kogut. (1999). "Localization of Knowledge and the Mobility of Engineers in Regional Networks." Management Science 45(7), 905-917.

Almeida, P., A. Phene and S. Li. (2014). "The Influence of Ethnic Community Knowledge on Indian Inventor Innovativeness." Organization Science 26(1), 198-217.

Amis, J.M., J. Mair and K.A. Munir. (2020). "The Organizational Reproduction of Inequality." Academy of Management Annals 14(1), 195-230.

Argyres, N.S. and B.S. Silverman. (2004). "R&D, Organization Structure, and the Development of Corporate Technological Knowledge." Strategic Management Journal 25, 929-958.

Arora, A., S. Belenzon and S. Lia. (2021a). "Knowledge Spillovers and Corporate Investment in Scientific Research." American Economic Review 111(3), 871-898.

Arora, A., S. Belenzon and S. Lia. (2021b). "Matching Patents to Compustat Firms, 1980-2015: Dynamic Reassignment, Name Changes, and Ownership Structures." Research Policy 50(5).

Arora, A., S. Belenzon and L.A. Rios. (2014). "Make, Buy, Organize: The Interplay between Research, External Knowledge, and Firm Structure." Strategic Management Journal 35(3), 317-337.

Audretsch, D. and M. Feldman. (1996). "R&D Spillovers and the Geography of Innovation and Production." American Economic Review 86, 630-640.

Bahar, D., P. Choudhury and B. Glennon. (2020). "An Executive Order Worth $100 Billion: The Impact of an Immigration Ban's Announcement on Fortune 500 Firms' Valuation." NBER Working Paper 27997.

Bahar, D. and H. Rapoport. (2018). "Migration, Knowledge Diffusion and the Comparative Advantage of Nations." Economic Journal 128(612), F273-F305.

Balachandran, S. and E. Hernandez. (2021). "Mi Casa Es Tu Casa: Immigrant Entrepreneurs as Pathways to Foreign Venture Capital Investments." Strategic Management Journal 42(11), 2047-2083.

Baldwin, C. and K. Clark. (2000). Design Rules: The Power of Modularity. MIT Press.

Bamford, C.E., T.J. Dean and P.P. McDougall. (2000). "An Examination of the Impact of Initial Founding Conditions and Decisions upon the Performance of New Bank Start-ups." Journal of Business Venturing 15(3), 253-277.

Baron, J.N., M.T. Hannan and M.D. Burton. (2001). "Labor Pains: Change in Organizational Models and Employee Turnover in Young, High-Tech Firms." American Journal of Sociology 106(4), 960-1012.

Bernstein, S., R. Diamond, T. McQuade and B. Pousada. (2022). "The Contribution of High-Skilled Immigrants to Innovation in the United States." NBER Working Paper 30797.

Bernstein, S., X. Giroud and R. Townsend. (2016). "The Impact of Venture Capital Monitoring." Journal of Finance 71(4), 1591-1622.

Boeker, W. (1989). "Strategic Change: The Effects of Founding and History." Academy of Management Journal 32(3), 489-515.

Branstetter, L., B. Glennon and J.B. Jensen. (2018). "The IT Revolution and the Globalization of R&D." NBER Innovation Policy and the Economy Volume 18, University of Chicago Press, edited by J. Lerner and S. Stern.

Breschi, S., F. Lissoni and E. Miguelez. (2017). "Foreign Inventors in the US: Testing for Diaspora and Brain Gain Effects." Journal of Human Resources 17(5), 1009-1038.

Burton, M.D. and C.M. Beckman. (2007). "Leaving a Legacy: Position Imprints and Successor Turnover in Young Firms." American Sociological Review 72(2), 239-266.

Byrne, D. (1971) The Attraction Paradigm. Academic Press, New York.

Carlino, G. and W.R. Kerr. (2015). "Agglomeration and Innovation." Handbook of Regional and Urban Economics. Volume 5, Elsevier, edited by G. Duranton, J. V. Henderson and W.C. Strange, 349-404.

Castellani, D., A. Perri and V.G. Scalera. (2022). "Knowledge Integration in Multinational Enterprises: The Role of Inventors Crossing National and Organizational Boundaries." Journal of World Business 57(3), 101290.

Choudhury, P. (2016). "Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants." Journal of Economic Geography 16(3), 585-610.

Choudhury, P. (2017). "Innovation Outcomes in a Distributed Organization: Intrafirm Mobility and Access to Resources." Organization Science 28(2), 339-354.

Choudhury, P. (2022). "Geographic Mobility, Immobility, and Geographic Flexibility—A Review and Agenda for Research on the Changing Geography of Work." Academy of Management Annals 16(1), 258-296.

Choudhury, P., T. Khanna and V. Sevcenko. (2022). "Firm-Induced Migration Paths and Strategic Human-Capital Outcomes." Management Science 69(1), 419-445.

Choudhury, P. and D.Y. Kim. (2019). "The Ethnic Migrant Inventor Effect: Codification and Recombination of Knowledge Across Borders." Strategic Management Journal 40(2), 203-229.

Cohen, W.M. and D.A. Levinthal. (1990). "Absorptive Capacity: A New Perspective on Learning and Innovation." Administrative Science Quarterly 35(1), 128-152.

Delgado, M. (2020). "The Co-location of Innovation and Production in Clusters." Industry and Innovation 27(8), 842-870.

Dimmock, S.G., J. Huang and S. Weisbenner. (2021). "Give Me Your Tired, Your Poor, Your High-Skilled Labor: H-1B Lottery Outcomes and Entrepreneurial Success." Management Science 68(9), 6950-6970.

Docquier, F. and H. Rapoport. (2012). "Globalization, Brain Drain, and Development." Journal of Economic Literature 50(3), 681-730.

Doran, K., A. Gelber and A. Isen. (2022). "The Effects of High-Skill Immigration on Firms: Evidence from H-1B Visa Lotteries." Journal of Political Economy 130(10), 2501-2533.

Duranton, G. and P.P. Combes. (2006). "Labour Pooling, Labour Poaching and Spatial Clustering." Regional Science and Urban Economics 36(1), 1-28.

Eklund, J. (2022). "The Knowledge-Incentive Tradeoff: Understanding the Relationship between Research and Development Decentralization and Innovation." Strategic Management Journal 43, 2478-2509.

Eklund, J. and R. Kappor. (2021). "Mind the Gaps: How Organization Design Shapes the Sourcing of Inventions." Organization Science 33(4), 1319-1339.

Ertug, G., J. Brennecke, B. Kovács and T. Zou. (2022). "What Does Homophily Do? A Review of the Consequences of Homophily." Academy of Management Annals 16(1), 38-69.

Ferrucci, E. and F. Lissoni. (2019). "Foreign Inventors in Europe and the United States: Diversity and Patent Quality." Research Policy 48(9), 103774.

Fleming, L., C. King and A. Juda. (2007). "Small Worlds and Regional Innovation." Organization Science 18(2), 938-954.

Foley, C.F. and W.R. Kerr. (2013). "Ethnic Innovation and U.S. Multinational Firm Activity." Management Science 59(7), 1529-1544.

Glennon, B. (2023). "How Do Restrictions on High-Skilled Immigration Affect Offshoring? Evidence from the H-1B Program." Management Science, forthcoming. NBER Working Paper 27538.

Hall, B., A. Jaffe and M. Trajtenberg. (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498 (2001).

Hannan, M.T. and J. Freeman. (1977). "The Population Ecology of Organizations." American Journal of Sociology 82(5), 929-964.

Hannan, M.T. and J. Freeman. (1984). "Structural Inertia and Organizational Change." American Sociological Review, 149-164.

Hegde, D. and J. Tumlinson. (2014). "Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity's Role in U.S. Venture Capital." Management Science 60(9), 2355-2380.

Hernandez, E. (2014) "Finding a Home Away From Home: Effects of Immigrants on Firms' Foreign Location Choice and Performance." Administrative Science Quarterly 59(1), 73-108.

Hernandez, E. and E. Kulchina. (2020). "Immigrants and Foreign Firm Performance." Organization Science 31(4), 797-820.

Kanter, R.M. (1977). Men and Women of the Corporation. New York: Basic Books.

Karim, S. and C. Williams. (2012). "Structural Knowledge: How Executive Experience with Structural Composition Affects Intra-firm Mobility and Unit Reconfiguration." Strategic Management Journal 33(6), 681-709.

Kerr, S.P. and W.R. Kerr. (2018). "Global Collaborative Patents." Economic Journal 128(612), F235-F272.

Kerr, S.P., W.R. Kerr and W.F. Lincoln. (2015). "Skilled Immigration and the Employment Structures of U.S. Firms." Journal of Labor Economics 33(S1), S147-S186.

Kerr, W.R. (2007). "The Ethnic Composition of U.S. Inventors." Harvard Business School Working Paper 08-006.

Kerr, W.R. (2019). The Gift of Global Talent: How Migration Shapes Business, Economy & Society. Stanford University Press, Palo Alto, CA.

Kerr, W.R. and W. Lincoln. (2010). "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention." Journal of Labor Economics 28(3), 473-508.

Kerr, W.R. and F. Robert-Nicoud. (2020). "Tech Clusters." Journal of Economic Perspectives 34(3), 50-76.

Kimberly, J.R. (1979). "Issues in the Creation of Organizations: Initiation, Innovation, and Institutionalization." Academy of Management Journal 22(3), 437-457.

Kleinbaum, A.M., T.E. Stuart and M.L. Tushman. (2013). "Discretion Within Constraint: Homophily and Structure in a Formal Organization." Organization Science 24(5), 1316-1336.

Kriauciunas, A. and P. Kale. (2006). "The Impact of Socialist Imprinting and Search on Resource Change: A Study of Firms in Lithuania." Strategic Management Journal 27(7), 659-679.

Lahiri, N. (2010). "Geographic Distribution of R&D Activity: How Does It Affect Innovation Quality?" Academy of Management Journal 53(5), 1194-1209.

Lawrence, P.R. and J.W. Lorsch. (1967). "Differentiation and Integration in Complex Organizations." Administrative Science Quarterly 12(1), 1-47.

Lawrence, B.S. and N.P. Shah. (2020). "Homophily: Measures and Meaning." Academy of Management Annals 14(2), 513-597.

Leiponen, A. and C.E. Helfat. (2011). "Location, Decentralization, and Knowledge Sources for Innovation." Organization Science 22(3), 641-658.

Lerner, J. and J. Wulf. (2007). "Innovation and Incentives: Evidence from Corporate R&D." The Review of Economics and Statistics 89(4), 634-644.

Li, G., R. Lai, D. Doolin, A. D'Amour, A. Yu, Y. Sun, V. Torvik and L. Fleming. (2014). "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database, 1975-2010." Research Policy 43, 941-955.

Madsen, T.L., E. Mosakowski and S. Zaheer. (2003). "Knowledge Retention and Personnel Mobility: The Nondisruptive Effects of Inflows of Experience." Organization Science 14(2), 173-191.

Marino, A., R. Mudambi, A. Perri and V.G. Scalera. (2020). "Ties That Bind: Ethnic Inventors in Multinational Enterprises' Knowledge Integration and Exploitation." Research Policy 49(9), 103956.

Marquis, C. and A. Tilcsik. (2013). "Imprinting: Toward a Multilevel Theory." Academy of Management Annals 7(1), 195-245.

Mayda, A.M., F. Ortega, G. Peri, K. Shih and C. Sparber. (2018). "The Effect of the H-1B Quota on the Employment and Selection of Foreign-Born Labor." European Economic Review 108, 105-128.

McPherson, J.M., L. Smith-Lovin and J.M. Cook. (2001). "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology 27(1), 415-444.

Miguelez, E. (2019). "Collaborative Patents and the Mobility of Knowledge Workers." Technovation 86-87, 62-74.

Miguelez, E. and C. Fink. (2017). "Measuring the International Mobility of Inventors: A New Database." In The International Mobility of Talent and Innovation - New Evidence and Policy Implications, Cambridge University Press.

Moretti, E. (2021). "The Effect of High-Tech Clusters on the Productivity of Top Inventors." American Economic Review 111(10), 3328-3375.

Nerkar, A. and S. Paruchuri. (2005). "Evolution of R&D Capabilities: The Role of Knowledge Networks Within a Firm." Management Science 15(5), 771-785.

Nguyen, K.T. (2020). "Trust and Innovation within the Firm: Evidence from Matched CEO-Firm Data." Working Paper.

Peri, G., K. Shih and C. Sparber. (2015). "STEM Workers, H-1B Visas, and Productivity in US Cities." Journal of Labor Economics 33(S1), S225-S255.

Png, I.P.L. (2019). "U.S. R&D, 1975–1998: A New Dataset." Strategic Management Journal 40, 715-735.

Rauch, J. (2001). "Business and Social Networks in International Trade." Journal of Economic Literature 39(4), 1177-1203.

Rivera, L.A. (2015). "Go With Your Gut: Emotion and Evaluation in Job Interviews." American Journal of Sociology 120(5), 1339-1389.

Ruiz Castro, M. and E. Holvino. (2016). "Applying Intersectionality in Organizations: Inequality Markers, Cultural Scripts and Advancement Practices in a Professional Service Firm." Gender, Work & Organization 23(3), 328-347.

Saxenian, A.L. (2006). The New Argonauts (Cambridge, MA: Harvard University Press).

Saxenian, A.L., with Y. Motoyama and X. Quan. (2002). Local and Global Networks of Immigrant Professionals in Silicon Valley (San Francisco, CA: Public Policy Institute of California).

Schein, E.H. (1983). "The Role of the Founder in Creating Organizational Culture." Organizational Dynamics 12(1), 13-28.

Shaver, J.M. and F. Flyer. (2000). "Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States." Strategic Management Review 21, 1175-1193.

Singh, J. (2005). "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." Management Science 51(5), 756-770.

Singh, J. (2008). "Distributed R&D, Cross-Regional Knowledge Integration and Quality of Innovative Output." Research Policy 37(1), 77-96.

Stinchcombe, A.L. (1965) "Social Structure and Organizations." In March, J.P., Ed., Handbook of Organizations, Rand McNally, Chicago, 142-193.

Stinchcombe, A.L. (2000). "Social Structure and Organizations." In Economics Meets Sociology in Strategic Management. Emerald Group Publishing Limited.

Tecu, I. (2013). "The Location of Industrial Innovation: Does Manufacturing Matter?" US Census Bureau Center for Economic Studies Paper No. CES-WP-13-09.

Tzabbar, D., B. Cirillo and S. Breschi. (2021). "The Differential Impact of Intrafirm Collaboration and Technological Network Centrality on Employees' Likelihood of Leaving the Firm." Organization Science 33(6), 2250-2273.

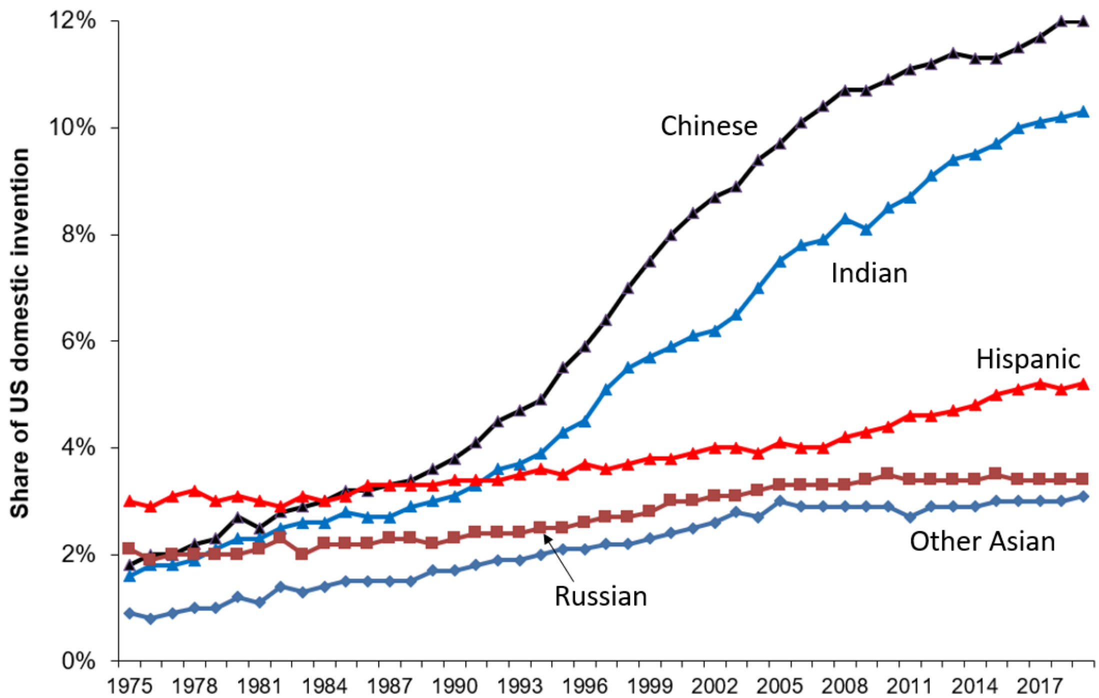
Waeger, D. and K. Weber. (2019). "Institutional Complexity and Organizational Change: An Open Polity Perspective." Academy of Management Review 44(2), 336-359.

Wang, D. (2015). "Activating Cross-border Brokerage: Interorganizational Knowledge Transfer through Skilled Return Migration." Administrative Science Quarterly 60(1), 133-176.

Williams, C.L., C. Muller and K. Kilanski. (2012). "Gendered Organizations in the New Economy." Gender & Society 26(4), 549-573.

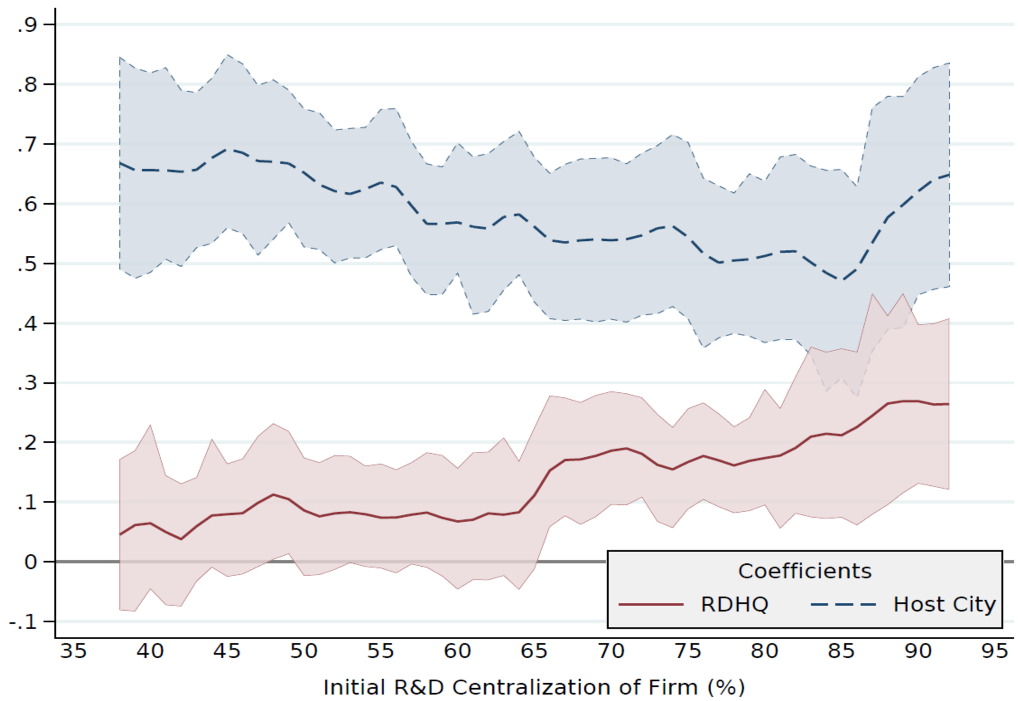
Wuchty, S., B.F. Jones and B. Uzzi. (2007). "The Increasing Dominance of Teams in Production of Knowledge." Science 316(5827), 1036-1039.

Figure 1: Ethnic composition of US invention



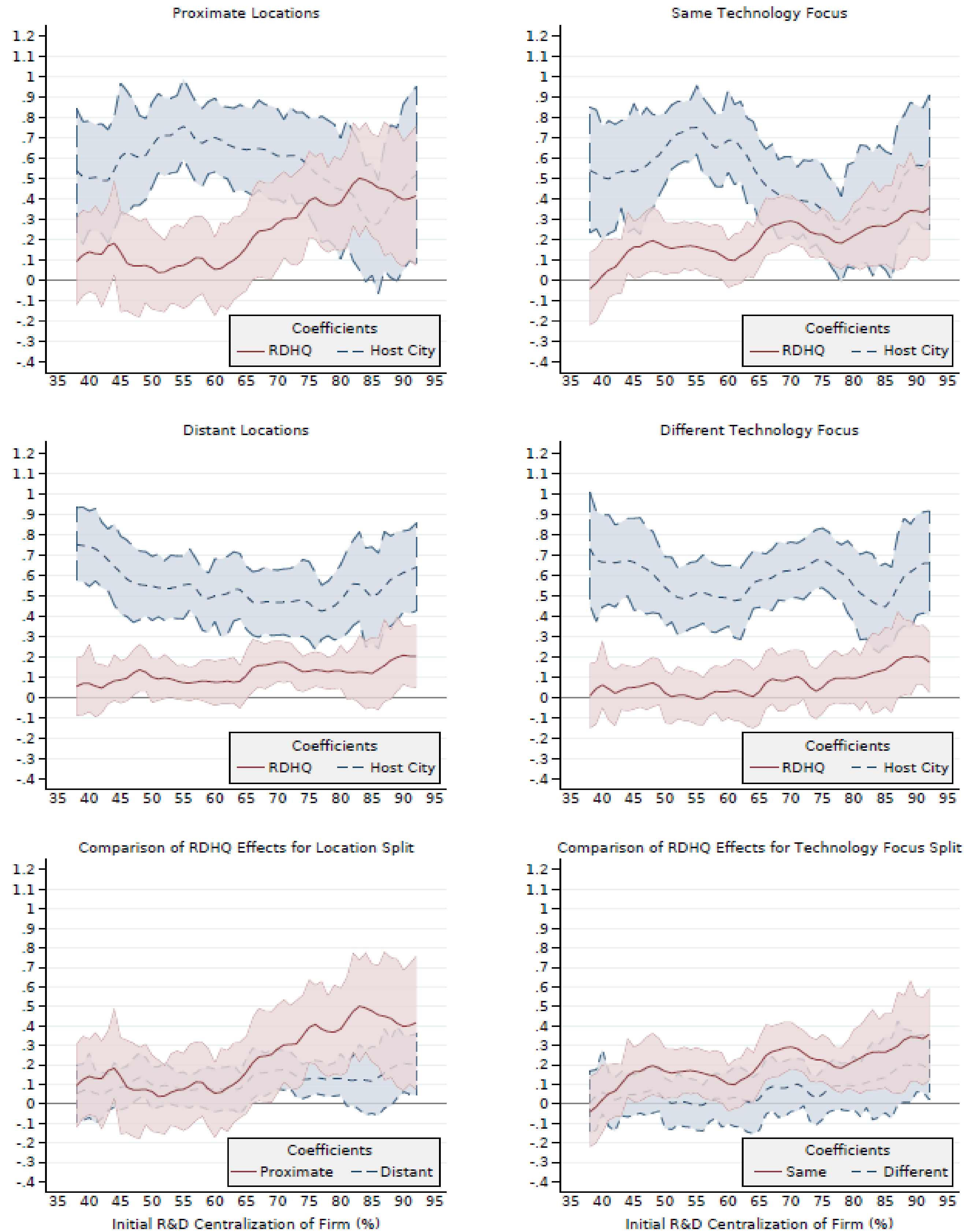
Notes: Figure presents the ethnic composition of US domestic inventors. Anglo-Saxon and European contributions (not shown) collectively decline from 90.6% in 1975 to 66.0% in 2019. Other Asian includes Japanese, Korean and Vietnamese.

Figure 2: Explanatory power for ethnic inventor composition at satellite locations



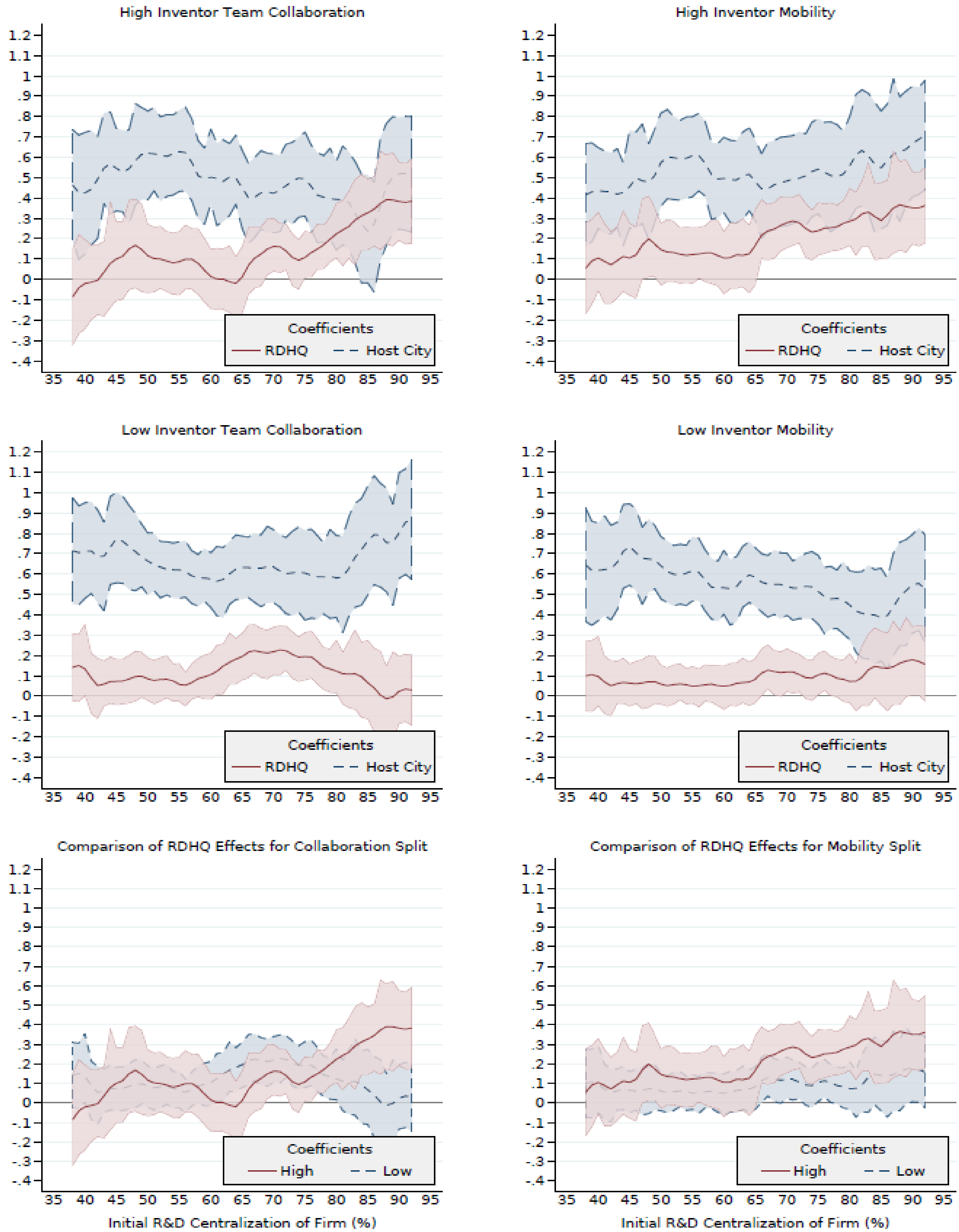
Notes: See Table 2 and Appendix Table 4. Figure presents outcomes from localized regressions that employ 16% bands on initial R&D centralization in the firm. For example, firms with initial R&D centralization of 30%-46% are used to compute the relationship shown at 38%.

Figure 3: Explanatory power by distance group and technology similarity



Notes: See Figure 2 and Table 2.

Figure 4: Explanatory power by interactions of satellite with RDHQ



Notes: See Figure 2 and Tables 4a and 4b.

Table 1: Descriptive statistics for full sample

	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)
Count of distinct satellite locations	7587			
Count of satellite locations - time periods	12222			
… Inventor-patent pairs	67.8	310.1	5	10821
… Ethnic inventor share	0.169	0.213	0.000	1.000
… Expected ethnic inventor share	0.214	0.117	0.000	1.000
… Female inventor share	0.071	0.138	0.000	1.000
… More than 1500 mi. from RDHQ	0.294	0.456	0.000	1.000
… Different NBER subcategory vs RDHQ	0.544	0.498	0.000	1.000
… Rate of collaborating with RDHQ	0.301	0.330	0.000	1.000
… Rate of intra-firm career mobility to RDHQ	0.110	0.219	0.000	1.000
Count of firms	444			
… Ethnic inventor share of RDHQ	0.236	0.138	0.000	0.843
… Expected ethnic inventor share of RDHQ	0.203	0.069	0.058	0.556
… Minimum RDHQ share of firm patenting	0.558	0.158	0.300	0.977
… Share in ICT fields	0.448	0.497	0.000	1.000
R&D centralization in initial period	0.647	0.172	0.316	0.983

Notes: Table presents descriptive statistics built on Discern-based parent firm sample. Sample includes firms with 250+ patents during 1975-2014 and US-based inventors accounting for 50%+ of firm's patents. RDHQ is the dominant R&D center of the firm. Patents are grouped into four decade-long time periods using application years: 1975-1984, 1985-1994, 1995-2004, and 2005-2014. Included firm-periods have 100+ inventor-patents. Satellite locations are defined at the MSA level and must contain 5+ inventor-patents in an included period. ICT firms are defined as those having a modal NBER category of patenting (across all facilities and locations) in Computers and Communications or Electrical and Electronic. R&D centralization of the firm is measured in the initial period of the firm as the share of the firm's US-based patents that occurred in RDHQ.

Table 2: Baseline estimations with full sample

	Baseline regression	Locations within 500 miles or direct plane flight of RDHQ	Locations NOT within 500 miles or direct plane flight of RDHQ	Locations with same tech focus as RDHQ	Locations with different tech focus than RDHQ	Separating firms by initial R&D centralization		
						(30%, 60%]	(60%, 75%]	(75%, 100%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DV is the ethnic share of inventors for the focal satellite R&D location							
Ethnic share of RDHQ	0.149+++	0.238+++	0.121+++	0.205+++	0.091++	0.066+	0.166+++	0.242+++
	(0.029)	(0.053)	(0.031)	(0.038)	(0.035)	(0.040)	(0.054)	(0.056)
Expected ethnic share for satellite location	0.612+++	0.615+++	0.600+++	0.544+++	0.651+++	0.643+++	0.552+++	0.590+++
	(0.037)	(0.056)	(0.048)	(0.056)	(0.047)	(0.048)	(0.067)	(0.079)
Expected ethnic share for RDHQ	0.107	0.044	0.165+	0.153	0.068	0.218+	0.039	0.035
	(0.090)	(0.141)	(0.099)	(0.121)	(0.109)	(0.129)	(0.189)	(0.210)
Observations	12222	4113	8109	5578	6644	5208	3731	3283
R-Squared value	0.199	0.246	0.200	0.226	0.216	0.239	0.239	0.215
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear difference of column pairs (2-3, 4-5)		0.117++		0.114+++				
		(0.057)		(0.044)				

Notes: See Table 1. Estimations consider the ethnic inventor share of satellite locations for a firm, with locations defined through peripheral MSAs and four time periods (1975-1984, 1985-1994, 1995-2004, 2005-2014). The primary regressor is the contemporaneous ethnic inventor share of RDHQ, reported in the first row. Estimations control for the expected ethnic invention shares of the satellite location and its RDHQ based upon contemporaneous invention happening in the same MSA and technology as the satellite location's workforce. Estimations control for MSA fixed effects, Tech-Period fixed effects, and log patent counts of the satellite location and its patent firm. Estimations are unweighted and cluster standard errors by parent firm. Columns 2 and 3 split the sample for satellite locations based upon being within 500 miles or a direct plane flight from RDHQ. Columns 4 and 5 split the sample for satellite locations based upon having the same technology focus as its RDHQ, defined using modal NBER subcategories. Columns 6-8 split the same by initial R&D centralization for the firm. The bottom row reports the linear difference between column pairs for the coefficients on the Ethnic share of RDHQ.

Table 3: Interaction estimations dividing sample by RDHQ involvement

		Rates of inventor collaboration on patents between location and RDHQ		Rates of inventor mobility during career between location and RDHQ	
	Baseline regression	Above median	Median and below	Above median	Median and below
	(1)	(2)	(3)	(4)	(5)
	DV is the ethnic share of inventors for the focal satellite R&D location				
Ethnic share of RDHQ	-0.072	-0.161	0.067	-0.106	-0.011
	(0.077)	(0.137)	(0.083)	(0.134)	(0.083)
x Initial R&D central.	0.325+++	0.491+++	0.065	0.484+++	0.155
	(0.108)	(0.183)	(0.114)	(0.174)	(0.121)
Expected ethnic share for satellite location	0.762+++	0.697+++	0.794+++	0.827+++	0.677+++
	(0.098)	(0.145)	(0.125)	(0.151)	(0.129)
x Initial R&D central.	-0.228	-0.276	-0.179	-0.406+	-0.149
	(0.145)	(0.212)	(0.177)	(0.214)	(0.177)
R&D centralization of firm in initial period	-0.029	-0.063	0.021	-0.029	-0.014
	(0.028)	(0.049)	(0.036)	(0.044)	(0.035)
Expected ethnic share for RDHQ	0.112	0.159	0.107	0.156	0.152
	(0.092)	(0.143)	(0.099)	(0.138)	(0.100)
Observations	12222	6062	6160	5109	7113
R-Squared value	0.200	0.198	0.236	0.261	0.176
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes
Linear combination:					
30% centralization	0.025	-0.014	0.086	0.039	0.036
60% centralization	0.123+++	0.134+++	0.106+++	0.184+++	0.082+++
90% centralization	0.221+++	0.281+++	0.125+++	0.329+++	0.129+++
Linear difference of column pairs (2-3, 4-5) for focal interaction		0.427++		0.329+	
		(0.210)		(0.185)	

Notes: See Table 2. Estimations incorporate interactions and a main effect for the spatial R&D centralization of the firm in the initial period, measured as the share of the firm's US-based patents that occurred in RDHQ. The bottom rows report combined effects at 30%, 60%, and 90% centralization. Columns 2 and 3 split the sample for satellite locations based upon their rate of co-invention with RDHQ. Columns 4 and 5 split the sample for satellite locations based upon rate of career mobility of inventors with RDHQ. The bottom row reports the linear difference between column pairs for the coefficients on the Ethnic share of RDHQ x Initial R&D centralization regressor.

Table 4a: R&D centralization and interactions with satellite location

	Rate of inventor collaboration over locations	Rate of inventor mobility over locations	Prior patenting experience of inventors in satellite location		
			Total	Movers with RDHQ	Local only
	(1)	(2)	(3)	(4)	(5)
R&D centralization of firm in initial period	0.384+++	0.192+++	4.119+	5.086	1.400
	(0.046)	(0.025)	(2.279)	(4.919)	(1.686)
Observations	12222	12222	12222	5109	11967
R-Squared value	0.191	0.092	0.145	0.140	0.158
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes

Notes: See Table 1. Estimations consider traits of satellite locations by the firm's initial level of R&D centralization. Estimations control for MSA fixed effects, Tech-Period fixed effects, and log patent counts of the satellite location and its patent firm. Estimations are unweighted and cluster standard errors by parent firm. Columns 1 and 2 consider rates of inventor collaboration and mobility, respectively. Columns 3-5 consider prior patenting experience of inventors who are present in the satellite location.

Table 4b: R&D centralization and patent citations made by satellite location

	Rate of internal-to-firm patent citations from satellite location	Rate of internal-to-RDHQ patent citations from satellite location	Rate of external-to-firm patent citations from satellite location	Share of patent citations that are internal-to-firm from satellite location	Share of patent citations that are to internal-to-RDHQ from satellite location
	(1)	(2)	(3)	(4)	(5)
R&D centralization of firm in initial period	0.211+	0.268+++	-0.050	0.051++	0.100+++
	(0.124)	(0.059)	(0.442)	(0.021)	(0.014)
Observations	12222	12222	12222	10931	10931
R-Squared value	0.051	0.036	0.094	0.133	0.071
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes

Notes: See Table 4a. Estimations consider rates and shares of patent citations made by the satellite location to prior work within the firm, prior work at RDHQ, and external prior work.

Table 5a: Interaction estimations by inventor type

	DV is ethnic share of inventors in satellite location with indicated trait				
	All inventors	Inventors collaborating on a patent with RDHQ	Inventors NOT collaborating on a patent with RDHQ	Inventors patenting at RDHQ at least once in career	Inventors NOT patenting at RDHQ at least once in career
	(1)	(2)	(3)	(4)	(5)
Ethnic share of RDHQ	-0.072	-0.051	-0.061	0.203	-0.057
	(0.077)	(0.126)	(0.077)	(0.242)	(0.073)
x Initial R&D central.	0.325+++	0.444+++	0.242++	0.460	0.257+++
	(0.108)	(0.158)	(0.111)	(0.302)	(0.100)
Expected ethnic share for satellite location	0.762+++ (0.098)	0.557+++ (0.147)	0.796+++ (0.100)	0.932+++ (0.236)	0.779+++ (0.097)
x Initial R&D central.	-0.228	-0.167	-0.198	-0.617+	-0.241+
	(0.145)	(0.217)	(0.151)	(0.330)	(0.141)
R&D centralization of firm in initial period	-0.029 (0.028)	-0.077 (0.049)	-0.012 (0.032)	-0.013 (0.080)	-0.018 (0.031)
Expected ethnic share for RDHQ	0.112 (0.092)	0.195 (0.131)	0.114 (0.088)	-0.212 (0.197)	0.149+ (0.090)
Observations	12222	8847	11361	5109	11967
R-Squared value	0.200	0.140	0.181	0.136	0.198
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes
Linear combination:					
30% centralization	0.025	0.083	0.012	0.341++	0.021
60% centralization	0.123+++	0.216+++	0.084+++	0.478+++	0.098+++
90% centralization	0.221+++	0.349+++	0.156+++	0.616+++	0.175+++

Notes: See Table 3.

Table 5b: Interaction estimations by inventor type, continued

	DV is ethnic share of inventors in satellite location with indicated trait				
	All inventors	Inventors collaborating on a patent with RDHQ	Inventors NOT collaborating on a patent with RDHQ	Inventors patenting at RDHQ at least once in career	Inventors NOT patenting at RDHQ at least once in career
	(1)	(2)	(3)	(4)	(5)
Ethnic share of RDHQ	-0.074	0.024	-0.165	0.179	-0.128
	(0.130)	(0.186)	(0.134)	(0.255)	(0.132)
x Initial R&D central.	0.380++	0.432+	0.371++	0.460	0.347++
	(0.163)	(0.226)	(0.172)	(0.308)	(0.175)
Expected ethnic share for satellite location	0.853+++	0.746+++	0.978+++	1.060+++	0.861+++
	(0.149)	(0.216)	(0.181)	(0.261)	(0.167)
x Initial R&D central.	-0.289	-0.322	-0.358	-0.566	-0.403
	(0.238)	(0.337)	(0.276)	(0.373)	(0.247)
R&D centralization of firm in initial period	-0.027	-0.041	0.002	-0.034	0.010
	(0.049)	(0.075)	(0.060)	(0.092)	(0.059)
Expected ethnic share for RDHQ	0.201	0.299+	0.163	-0.138	0.309++
	(0.124)	(0.155)	(0.145)	(0.216)	(0.136)
Observations	4313	4313	4313	4313	4313
R-Squared value	0.302	0.201	0.245	0.144	0.272
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes
Linear combination:					
30% centralization	0.040	0.153	-0.054	0.318+	-0.023
60% centralization	0.154+++	0.283+++	0.058	0.456+++	0.081+
90% centralization	0.268+++	0.413+++	0.169+++	0.594+++	0.185+++

Notes: See Table 5a. The sample is restricted to observations with all four types of inventors present.

Table 6: Persistence of ethnic invention at RDHQ for legacy sample

	DV is ethnic share of inventors in RDHQ with indicated trait:				
	All inventors in indicated period	Inventors who first patented with firm in 1975-1984	Inventors who first patented with firm in 1985-1994	Inventors who first patented with firm in 1995-2004	Inventors who first patented with firm in 2005-2014
	(1)	(2)	(3)	(4)	(5)
	A. Ethnic inventors patenting during 1995-2004				
Ethnic share of RDHQ during 1975-1984	0.729+++	1.050+++	0.678+++	0.659+++	n.a.
	(0.136)	(0.244)	(0.121)	(0.164)	
Expected ethnic share for RDHQ in period	0.477++	-0.543	0.653+++	0.502	
	(0.218)	(0.466)	(0.247)	(0.324)	
Observations	126	124	125	126	
R-Squared value	0.435	0.286	0.328	0.303	
Tech-Pd FE	Yes	Yes	Yes	Yes	
	B. Ethnic inventors patenting during 2005-2014				
Ethnic share of RDHQ during 1975-1984	0.627+++	0.893++	0.883+++	0.527++	0.366+
	(0.161)	(0.423)	(0.268)	(0.204)	(0.217)
Expected ethnic share for RDHQ in period	1.133+++	0.678	0.650	1.085+++	1.311+++
	(0.343)	(0.946)	(0.411)	(0.383)	(0.439)
Observations	83	73	80	82	83
R-Squared value	0.480	0.225	0.292	0.344	0.363
Tech-Pd FE	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2. Table considers ethnic workforce composition over time at RDHQ of firms patenting after 1995 and their ethnic invention during 1975-1984.

Table 7: Persistence estimations for satellite locations in legacy sample

	Baseline regression	Locations within 500 miles or direct plane flight of RDHQ	Locations NOT within 500 miles or direct plane flight of RDHQ	Locations with same tech focus as RDHQ	Locations with different tech focus than RDHQ	Separating firms by initial R&D centralization		
						(30%, 60%]	(60%, 75%]	(75%, 100%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DV is the ethnic share of inventors for the focal satellite R&D location							
Ethnic share of RDHQ during 1975-1984	0.200+++	0.238++	0.161++	0.221++	0.181++	-0.101	0.175+	0.370++
	(0.057)	(0.099)	(0.062)	(0.091)	(0.074)	(0.077)	(0.103)	(0.178)
Expected ethnic share for satellite location	0.664+++	0.753+++	0.582+++	0.592+++	0.723+++	0.730+++	0.600+++	0.693+++
	(0.059)	(0.084)	(0.099)	(0.101)	(0.077)	(0.072)	(0.106)	(0.119)
Expected ethnic share for RDHQ	0.069	0.026	0.073	0.090	0.051	0.579++	-0.012	0.038
	(0.126)	(0.205)	(0.147)	(0.165)	(0.175)	(0.217)	(0.295)	(0.258)
Observations	4142	1491	2651	1676	2466	1522	1576	1044
R-Squared value	0.222	0.299	0.238	0.266	0.276	0.334	0.317	0.277
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear difference of column pairs (2-3, 4-5)		0.077		0.040				
		(0.114)		(0.122)				

Notes: See Tables 2 and 6. Table considers ethnic invention in satellite locations of firms in operation after 1995 with the ethnic workforce present in RDHQ during 1975-1984.

Appendix to
Centralization and Organization Reproduction: Ethnic Innovation in R&D Centers and Satellite Locations

October 2023

1 Data Preparation

We downloaded patents through the PatentsView.org web API. We focus on patents applied for during 1975-2014. Application years provide the best timing of inventive work given the lengthy review process. With data through year end 2020, we are ensured that almost all patents applied for in 2014 and earlier have completed the USPTO review process. As few patents complete the process in one year, we also have a full sample for 1975.

The USPTO used the United States Patent Classification (USPC) system for classifying the technologies of granted patents from the start of our sample until it moved to the Cooperative Patent Classification (CPC) system in 2013. Patent classes from the USPC system were the foundation for the NBER Category system (Hall et al. 2001). To extend the NBER system to the end of our data, we developed a probabilistic NBER-CPC mapping based upon the transition period during which the USPTO assigned both codes. In the NBER system, there are 6 main categories and 36 subcategories.

The longitudinal tracking of inventors across patents is important. Li et al. (2014) pioneered name disambiguation techniques to link inventor names on patents into longitudinal inventor identifiers, correcting for minor spelling differences or nicknames. PatentsView provides inventor IDs based upon the USPTO's continuation of this important effort. We utilize these inventor IDs, with a modest set of modifications that improve upon the algorithms for combining inventor IDs within a company that are very likely the same individual.

Our most significant intervention is to correct cases where the names of inventors were parsed differently. For example, one record might have {"Mark", "Anthony", "Hernandez"} for the {first, middle, last} name fields, while another record had {"Mark Anthony", "", "Hernandez"} with no middle name. The PatentsView algorithm rarely recognizes these to be the same individual. We combine the inventor IDs of PatentsView together in situations like these where the records were very likely the same person, such as having contemporaneous patents with the same company, geographic location, and technology domain. The modifications are available to other researchers in the replication data of this paper.

PatentsView provides a binary gender estimate based upon name algorithms. There is a particular challenge for isolating gender with Asian names, which held the potential to bias our work. We thus manually examined the most prominent inventors for accuracy, including contracting with freelancers to web search over 2600 prolific inventors we could not initially confirm. These gender assignments are included in the data replication package for this paper.

Assignment of ethnicity uses first, middle, and last names of individuals, and the algorithms assign a probability across multiple ethnicities where the name is ambiguous (e.g., the surname "Lee" is linked to Anglo-Saxon, Chinese and Korean ethnicities). In cases where an individual has a multi-ethnic mapping, such as 50% Chinese and 50% Anglo-Saxon, we use the partial ethnic shares in location averages.

Ethnicity is imperfectly correlated with migration—e.g., some Anglo-Saxon inventors are recent migrants from the UK, while some Chinese inventors are second-generation citizens. The procedure also does not capture name changes with marriage or to Anglicize first names. Yet, most of the growth in ethnic invention during 1975-2014, and especially Chinese and Indian patenting, is linked to migration (e.g., Kerr and Lincoln 2010). One also observes comparable trends with the Census of Populations for the role of first-generation immigrants engaged in science and engineering. The micro-level assignment of ethnicity allows exploration within firms of their inventive workforces.

Our graphs that compare levels of centralization are built as follows: Our estimation design requires RDHQ hold at least 30% of the firm's patents, and we have a max share of 98.3%. For each increment on the horizontal axis, we extract a subsample that is within an 8% window in R&D centralization. This procedure develops 55 subsamples starting from (30%, 46%] for the sample centered on 38% and ending with [84%,100%) for the sample centered on 92%. We finally estimate specification (1) for each of the 55 subsamples and plot the non-parametric series of β and γ_1 coefficients.

2 Additional Results

Appendix Table 2a shows robustness to specification (1) when including fixed effects for deciles of patent counts, excluding control variables, weighting locations by their patent counts, weighting locations using Coarsened Exact Matching[1], winsorizing ethnic shares at their 95th percentile values, and using five-year time periods. Results are very similar when replacing $\eta_{c_f,t}$ with ethnic shares in USPC classes by period.

Appendix Table 2b shows additional variations. Our core estimation includes fixed effects

[1] In preparation for later study of R&D centralization, we design matching to achieve comparability of firms above and below median R&D centralization. Included variables in the matching are RDHQ ethnicity, inventor-patent counts of satellite location, inventor-patent counts of parent firm, and technology-period.

for satellite MSAs to broadly control for local ethnic workforces; by contrast, it does not include fixed effects for RDHQ locations to allow for more between-firm variation. Column 1 shows results when excluding satellite fixed effects; Column 2 includes fixed effects for satellite MSAs and for RDHQ locations. The results are very consistent across specifications. In the preferred specification, we estimate that RDHQ can account for an effect on satellite workforce equal to 24% of the satellite's host city effect. In these variations, we observe a range in equivalent explanatory power of 19% (Column 1) to 27% (Column 2).

Our baseline specification is unweighted and with a unit of observation at the satellite-period level. This approach implicitly provides more weight to firms with more satellites. Column 3 shows very similar results when we weight observations such that firms carry equal weight in each period.

There could be some concern that satellite facilities are constrained by competition for talent from other firms. To consider this, Column 4 introduces a control for the total count of observations in an MSA-technology cell. This specification and variants like it suggest the effect is not influenced by the size of the local cluster. Column 5 alternatively models the size rank of the satellite in the parent firm's portfolio, again finding very consistent results. Additionally, the patterns are similar in and out of tech hubs. The evidence for organization reproduction is modestly stronger when RDHQs are more alike to their home city than when they are already different.

Our baseline approach treats all inventor-patent observations equally when determining ethnicity, which provides more implicit weight to patents that have more inventors. For Column 6, we recompute ethnic shares so that all patents have the same emphasis, downweighting inventor records when there are more inventors on the patent. This last column shows similar outcomes.

We also find comparable results in Appendix Table 2c when only including satellite locations with >20 or >50 inventor-patents and when only including firms that have a single RDHQ that always comprises >50% of domestic patenting. The role of spatial centralization also holds when splitting the sample by the assignee filing decentralization measure developed by Arora et al. (2014). We finally observe similar results when conducting our analysis with all corporate assignees (public and private companies) in the PatentsView data.

Appendix Table 2d shows additional breakouts. We control in specification (1) for the expected inventor ethnicities of the satellite and RDHQ, given their locations and technology focus, to guard against spurious correlations due to trends in sector growth or ethnic patenting growth. The first four columns consider two sample splits for additional robustness. We calculate for each patent sub-category 1) the total growth of invention from the first period of 1975-1984 to the final period of 2005-2014 and 2) the total growth of ethnic invention from 1975-1984 to 2005-2014. Mapping these technology growth rates to each satellite based upon the satellite's core technology focus, we find quite similar results in technologies above and below the median growth rates. The last two columns also show similar patterns when splitting the sample based upon parent firm diversification, measured by the number of distinct sub-category technologies being patented by the firm in satellites. We find similar results above and below median. While we acknowledge our inability to measure the strategy of firms with respect to their satellites, this split suggests the outcomes are not a product of different approaches being pursued by diversified R&D firms.

Appendix Tables 2e-8 are noted in the text.

Finally, the Discussion and Future Research section notes the importance of assessing the implications of these choices for firms. An organization may aspire to access new talent and expand its innovation scope (e.g., Kogut and Zander 1992), but its success requires balancing leveraging its existing networks with the construction of new operations attuned to a desired location (e.g., Romanelli and Khessina 2005; Alcacer and Chung 2007; Laursen et al. 2011). Stronger internal bonds may boost knowledge flows within the firm, but they may also weaken the inflow of external information into satellite facilities. While studies discussed in Section 2 tend to find centralized systems deliver more impactful innovations, it is not clear whether this applies to satellite locations of these firms.

In Appendix Tables 9 and 10, we provide some initial exploration using patent traits. Compared to the patents of their parent RDHQ, the patents of satellite locations tend to be less original and have smaller stock market reactions on announcement (Kogan et al. 2017). These periphery patents also have a future citation pattern that is more external than internal to the firm. These results may indicate a more development focus of R&D work in satellite locations. When comparing among satellite locations of a firm, satellites with links to RDHQ show higher patent novelty, stock market response to patent announcements, and internal citations.

Additional References

Alcacer, J. and W. Chung. (2007). "Location Strategies and Knowledge Spillovers." Management Science 53(5), 760-776.

Kogan, L., D. Papanikolaou, A. Seru and N. Stoffman. (2017). "Technological Innovation, Resource Allocation, and Growth." The Quarterly Journal of Economics 132(2), 665-712.

Kogut, B. and U. Zander. (1992). "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." Organization Science 3(3), 383-397.

Laursen, K., F. Masciarelli and A. Prencipe. (2011). "Regions Matter: How Localized Social Capital Affects Innovation and External Knowledge Acquisition." Organization Science 23(1), 177-193.

Romanelli, E. and O.M. Khessina. (2005). "Regional Industrial Identity: Cluster Configurations and Economic Development." Organization Science 16(4), 344-358.

Appendix Table 1a: Ethnic composition of inventors residing in United States

	Ethnicity of Inventor								
	Anglo-Saxon	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnam.
1975-1979	74.6	2.0	15.5	3.0	1.9	0.5	0.3	2.0	0.1
1980-1984	73.2	2.8	15.1	3.0	2.5	0.7	0.5	2.1	0.1
1985-1989	72.1	3.3	14.7	3.3	2.8	0.8	0.5	2.2	0.2
1990-1994	70.1	4.4	14.1	3.5	3.5	0.9	0.6	2.4	0.3
1995-1999	66.3	6.5	13.6	3.6	5.1	1.0	0.7	2.7	0.5
2000-2004	62.4	8.7	13.0	3.9	6.3	1.1	0.9	3.1	0.6
2005-2009	59.1	10.3	12.3	4.1	7.9	1.2	1.2	3.3	0.6
2010-2014	57.1	11.2	11.7	4.6	9.1	1.0	1.3	3.4	0.6
2015-2019	55.4	11.6	11.5	5.1	10.0	1.0	1.4	3.4	0.6
Chemicals	63.6	8.3	13.9	3.7	5.5	1.0	0.9	2.9	0.4
Computers	55.2	11.2	11.7	4.2	11.4	1.1	1.1	3.4	0.6
Pharmaceuticals	61.5	8.6	14.0	4.6	5.7	1.0	1.0	3.1	0.5
Electrical	59.2	11.0	12.6	3.8	6.7	1.2	1.4	3.4	0.6
Mechanical	71.6	3.9	13.8	3.6	3.0	0.8	0.5	2.5	0.3
Miscellaneous	73.1	3.2	13.4	4.2	2.5	0.6	0.5	2.2	0.3

Notes: Table presents descriptive statistics for inventors residing in the US at the time of patent application. Inventor ethnicities are estimated through inventors' names using techniques described in the text. Patents are grouped by application years and major technology fields.

Appendix Table 1b: Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Ethnic share of satellite	1.000					
(2) Ethnic share of RDHQ	0.2238*	1.000				
(3) Exp. ethnic share of satellite	0.4027*	0.3819*	1.000			
(4) Exp. ethnic share of RDHQ	0.2503*	0.6252*	0.6114*	1.000		
(5) Initial R&D centralization	0.0542*	0.1274*	0.1224*	0.2063*	1.000	
(6) Average R&D centralization	0.0631*	0.1090*	0.1125*	0.1878*	0.7415*	1.000

Notes: See Tables 1 and 2.

Appendix Table 2a: Baseline estimations with robustness checks

	Including fixed effects for deciles of firm patent counts	Excluding all control variables	Weighting locations by patent counts	Weighting locations by CEM split on median R&D conc.	Winsorized shares at 95th percentile	Utilizing 5-year time periods
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	0.146+++	0.171+++	0.185+++	0.205+++	0.144+++	0.145+++
	(0.029)	(0.026)	(0.046)	(0.033)	(0.026)	(0.026)
Expected ethnic share for satellite location	0.613+++	0.730+++	0.883+++	0.543+++	0.538+++	0.626+++
	(0.037)	(0.025)	(0.045)	(0.047)	(0.031)	(0.038)
Expected ethnic share for RDHQ	0.127	-0.194+++	-0.048	0.148	0.060	0.110
	(0.090)	(0.053)	(0.109)	(0.101)	(0.072)	(0.079)
Observations	12222	12222	12222	9177	12222	14101
R-Squared value	0.200	0.170	0.505	0.192	0.221	0.213
MSA and Tech-Pd FE	Yes	No	Yes	Yes	Yes	Yes
Size controls	Yes	No	Yes	Yes	Yes	Yes

Notes: See Table 2.

Appendix Table 2b: Baseline estimations with robustness checks, continued

	Including MSA fixed effects for RDHQ	Dropping all MSA fixed effects (RDHQ and satellite)	Weighting each firm equally (vs. satellites)	Controlling for location x period cluster size	Controlling for size rank of satellite in firm portfolio	Calculating ethnicity with equal weights to patents
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	0.164+++	0.142+++	0.203+++	0.149+++	0.148+++	0.169+++
	(0.032)	(0.029)	(0.059)	(0.029)	(0.029)	(0.030)
Expected ethnic share for satellite location	0.602+++	0.732+++	0.576+++	0.608+++	0.613+++	0.624+++
	(0.038)	(0.026)	(0.048)	(0.037)	(0.037)	(0.038)
Expected ethnic share for RDHQ	0.110	0.043	0.063	0.105	0.086	0.090
	(0.096)	(0.089)	(0.180)	(0.090)	(0.091)	(0.091)
Observations	12222	12222	12222	12222	12222	12222
R-Squared value	0.209	0.174	0.198	0.199	0.199	0.192
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2.

Appendix Table 2c: Baseline estimations with sample splits

	Locations with >20 patent-inventors	Locations with >50 patent-inventors	Firms with single RDHQ that is >50% of patents	Firms with above median assignee filing decentr.	Firms with median and below assignee filing decentr.	Utilizing PatentsView assignee sample
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	0.148+++	0.121+++	0.163+++	0.137+++	0.173+++	0.197+++
	(0.031)	(0.038)	(0.045)	(0.043)	(0.040)	(0.025)
Expected ethnic share for satellite location	0.778+++	0.885+++	0.572+++	0.606+++	0.604+++	0.684+++
	(0.059)	(0.069)	(0.063)	(0.050)	(0.053)	(0.032)
Expected ethnic share for RDHQ	0.147	0.255+	0.207	0.162	0.113	0.045
	(0.099)	(0.139)	(0.163)	(0.119)	(0.147)	(0.067)
Observations	4715	2411	5080	6050	6171	17019
R-Squared value	0.370	0.502	0.211	0.211	0.224	0.197
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2.

Appendix Table 2d: Baseline estimations with sample splits, continued

	Splitting the sample of technologies by the national rate of patenting growth among all inventors		Splitting the sample of technologies by the national rate of patenting growth among ethnic inventors		Splitting the sample of firms by their total span of distinct technologies across satellites	
	Median and above	Below median	Median and above	Below median	Median and above	Below median
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	0.162+++	0.137+++	0.161+++	0.138+++	0.105++	0.184+++
	(0.036)	(0.043)	(0.036)	(0.043)	(0.043)	(0.040)
Expected ethnic share for satellite location	0.593+++	0.599+++	0.608+++	0.587+++	0.662+++	0.506+++
	(0.058)	(0.057)	(0.057)	(0.057)	(0.049)	(0.055)
Expected ethnic share for RDHQ	0.042	0.221+	0.037	0.241+	0.160	0.172
	(0.124)	(0.124)	(0.124)	(0.123)	(0.142)	(0.124)
Observations	6199	6023	6170	6052	7415	4807
R-Squared value	0.221	0.188	0.222	0.185	0.227	0.207
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2.

Appendix Table 2e: Baseline estimations with additional distance variation

	Coefficients from distance band estimations		Column 1 with separation for direct flights
	(1)		(2)
Satellite locations within 100 mi. of RDHQ	0.337+++ (0.095)	Satellite locations within 100 mi. of parent firm's RDHQ	0.335+++ (0.096)
Satellite locations 100-500 miles of RDHQ	0.209+++ (0.047)	Satellite locations 100-500 miles of RDHQ	
		... With direct plane flight	0.219+++ (0.050)
		... Without direct plane flight	0.153++ (0.067)
Satellite locations 500-1500 miles of RDHQ	0.178+++ (0.034)	Satellite locations 500-1500 miles of RDHQ	
		... With direct plane flight	0.188+++ (0.039)
		... Without direct plane flight	0.153+++ (0.048)
Satellite locations 1500+ miles of RDHQ	0.156+++ (0.034)	Satellite locations 1500+ miles of RDHQ	
		... With direct plane flight	0.177+++ (0.041)
		... Without direct plane flight	0.116+++ (0.045)

Notes: See Table 2. Estimations report coefficients on ethnic share of RDHQ for satellite locations separated by distance bands and the presence of direct flights. Estimations include expected ethnic inventor shares at the satellite location and RDHQ, satellite and firm size controls, MSA and Tech-Year fixed effects, and fixed effects for distance band. Column 2 further includes a indicator variable for a direct flight.

Appendix Table 3: Baseline estimations with demographic compositions

	DV is ethnic share of inventors in satellite location	DV is Chinese and Indian ethnic share of inventors in satellite location	DV is non-Chinese and non-Indian ethnic share of inventors in satellite location	DV is ethnic share of inventors in satellite location	DV is female share of inventors in satellite location
	(1)	(2)	(3)	(4)	(5)
Ethnic share of RDHQ	0.149+++			0.151+++	-0.042++
	(0.029)			(0.030)	(0.017)
Expected ethnic share for satellite location	0.612+++	0.406+++	0.206+++	0.617+++	0.051++
	(0.037)	(0.028)	(0.027)	(0.037)	(0.021)
Expected ethnic share for RDHQ	0.107	0.123+	-0.020	0.120	0.104+
	(0.090)	(0.073)	(0.051)	(0.090)	(0.058)
Chinese and Indian ethnic share of RDHQ		0.182+++	-0.020		
		(0.030)	(0.018)		
Non-Chinese and Indian ethnic share of RDHQ		0.006	0.101+++		
		(0.047)	(0.038)		
Female inventor share of RDHQ				-0.029	0.324+++
				(0.062)	(0.045)
Expected female share for satellite location				-0.040	0.736+++
				(0.049)	(0.065)
Observations	12222	12222	12222	12222	12222
R-Squared value	0.199	0.162	0.093	0.199	0.150
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2.

Appendix Table 4: Baseline estimations by initial R&D centralization in firm

	Baseline regression	Separating firms by initial R&D centralization (30%, 45%]	(45%, 60%]	(60%, 75%]	(75%, 100%)
	(1)	(2)	(3)	(4)	(5)
	DV is the ethnic share of inventors for the focal satellite R&D location				
Ethnic share of RDHQ	0.149+++	-0.012	0.088+	0.166+++	0.242+++
	(0.029)	(0.062)	(0.046)	(0.054)	(0.056)
Expected ethnic share for satellite location	0.612+++	0.655+++	0.613+++	0.552+++	0.590+++
	(0.037)	(0.096)	(0.056)	(0.067)	(0.079)
Expected ethnic share for RDHQ	0.107	0.178	0.187	0.039	0.035
	(0.090)	(0.193)	(0.150)	(0.189)	(0.210)
Observations	12222	1683	3525	3731	3283
R-Squared value	0.199	0.361	0.242	0.239	0.215
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2 and Figure 2.

App. Table 5a: Interaction estimations with distance and technology splits

	Baseline regression	Locations within 500 miles or direct plane flight of RDHQ	Locations NOT within 500 miles or direct plane flight of RDHQ	Locations with same tech focus as RDHQ	Locations with different tech focus than RDHQ
	(1)	(2)	(3)	(4)	(5)
	DV is the ethnic share of inventors for the focal satellite R&D location				
Ethnic share of RDHQ	-0.072	-0.199	-0.063	-0.193+	-0.004
	(0.077)	(0.173)	(0.083)	(0.107)	(0.101)
x Initial R&D central.	0.325+++	0.650++	0.271++	0.583+++	0.141
	(0.108)	(0.259)	(0.115)	(0.159)	(0.137)
Expected ethnic share for satellite location	0.762+++	0.758+++	0.773+++	0.831+++	0.714+++
	(0.098)	(0.160)	(0.119)	(0.125)	(0.135)
x Initial R&D central.	-0.228	-0.222	-0.263	-0.438++	-0.096
	(0.145)	(0.234)	(0.177)	(0.190)	(0.191)
R&D centralization of firm in initial period	-0.029	-0.059	-0.016	-0.036	-0.017
	(0.028)	(0.045)	(0.037)	(0.041)	(0.041)
Expected ethnic share for RDHQ	0.112	-0.001	0.182+	0.162	0.072
	(0.092)	(0.141)	(0.100)	(0.123)	(0.113)
Observations	12222	4113	8109	5578	6644
R-Squared value	0.200	0.250	0.201	0.229	0.216
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes
<u>Linear combination:</u>					
30% centralization	0.025	-0.004	0.019	-0.018	0.039
60% centralization	0.123+++	0.191+++	0.100+++	0.157+++	0.081++
90% centralization	0.221+++	0.386+++	0.181+++	0.332+++	0.123+++
Linear difference of column pairs (2-3, 4-5) for focal interaction		0.379		0.442++	
		(0.269)		(0.185)	

Notes: See Tables 2 and 3.

App. Table 5b: App. Table 5a using lagged centralization values

	Baseline regression	Locations within 500 miles or direct plane flight of RDHQ	Locations NOT within 500 miles or direct plane flight of RDHQ	Locations with same tech focus as RDHQ	Locations with different tech focus than RDHQ
	(1)	(2)	(3)	(4)	(5)
	DV is the ethnic share of inventors for the focal satellite R&D location				
Ethnic share of RDHQ	-0.048	-0.441++	0.031	-0.084	-0.011
	(0.089)	(0.209)	(0.104)	(0.116)	(0.115)
x Initial R&D central.	0.298++	1.069+++	0.126	0.444+++	0.152
	(0.120)	(0.326)	(0.141)	(0.158)	(0.160)
Expected ethnic share for satellite location	0.689+++	0.818+++	0.674+++	0.720+++	0.681+++
	(0.101)	(0.167)	(0.130)	(0.144)	(0.127)
x Initial R&D central.	-0.123	-0.321	-0.120	-0.226	-0.076
	(0.161)	(0.269)	(0.198)	(0.228)	(0.198)
R&D centralization of firm in initial period	-0.042	-0.119++	-0.007	-0.036	-0.026
	(0.034)	(0.058)	(0.045)	(0.051)	(0.049)
Expected ethnic share for RDHQ	0.024	-0.100	0.092	0.085	-0.062
	(0.105)	(0.163)	(0.117)	(0.133)	(0.137)
Observations	9187	3080	6107	4131	5056
R-Squared value	0.202	0.265	0.202	0.241	0.221
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes
Linear combination:					
30% centralization	0.041	-0.120	0.069	0.049	0.035
60% centralization	0.130+++	0.201+++	0.107+++	0.182+++	0.080+
90% centralization	0.220+++	0.521+++	0.145+++	0.316+++	0.126++
Linear difference of column pairs (2-3, 4-5) for focal interaction		0.943++		0.292	
		(0.366)		(0.202)	

Notes: See Tables 2 and 3.

Appendix Table 6a: Interaction estimations with robustness checks

	Including fixed effects for deciles of firm patent counts	Excluding all control variables	Weighting locations by patent counts	Weighting locations by CEM split on median R&D conc.	Winsorized shares at 95th percentile	Utilizing 5-year time periods
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	-0.078	-0.013	-0.055	-0.076	-0.028	-0.033
	(0.078)	(0.081)	(0.093)	(0.119)	(0.069)	(0.073)
x Initial R&D central.	0.330+++	0.273++	0.348+++	0.418++	0.261+++	0.247++
	(0.112)	(0.113)	(0.121)	(0.167)	(0.095)	(0.098)
Expected ethnic share for satellite location	0.770+++	0.829+++	1.016+++	0.821+++	0.706+++	0.667+++
	(0.098)	(0.089)	(0.109)	(0.144)	(0.080)	(0.105)
x Initial R&D central.	-0.238	-0.151	-0.215	-0.422++	-0.255++	-0.061
	(0.144)	(0.138)	(0.192)	(0.194)	(0.119)	(0.152)
R&D centralization of firm in initial period	-0.022	-0.030	-0.054	-0.005	-0.011	-0.041
	(0.029)	(0.025)	(0.052)	(0.036)	(0.026)	(0.033)
Expected ethnic share for RDHQ	0.125	-0.197+++	-0.005	0.152	0.070	0.103
	(0.090)	(0.053)	(0.102)	(0.105)	(0.075)	(0.079)
Observations	12222	12222	12222	9177	12222	14101
R-Squared value	0.201	0.171	0.507	0.194	0.222	0.213
MSA and Tech-Pd FE	Yes	No	Yes	Yes	Yes	Yes
Size controls	Yes	No	Yes	Yes	Yes	Yes
<u>Linear combination:</u>						
30% centralization	0.021	0.069	0.050	0.049	0.050	0.037
60% centralization	0.120+++	0.151+++	0.154+++	0.175+++	0.128+++	0.114+++
90% centralization	0.219+++	0.233+++	0.259+++	0.300+++	0.206+++	0.191+++

Notes: See Table 3.

Appendix Table 6b: Interaction estimations with robustness checks, continued

	Including MSA fixed effects for RDHQ	Dropping all MSA fixed effects (RDHQ and satellite)	Weighting each firm equally (vs. satellites)	Controlling for location x period cluster size	Controlling for size rank of satellite in firm portfolio	Calculating ethnicity with equal weights to patents
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	-0.113	-0.080	-0.248+	-0.076	-0.076	-0.086
	(0.079)	(0.076)	(0.143)	(0.076)	(0.077)	(0.080)
x Initial R&D central.	0.405+++	0.327+++	0.635+++	0.323+++	0.319+++	0.375+++
	(0.116)	(0.107)	(0.244)	(0.106)	(0.109)	(0.114)
Expected ethnic share for satellite location	0.753+++	0.881+++	0.951+++	0.854+++	0.882+++	0.784+++
	(0.099)	(0.092)	(0.146)	(0.093)	(0.092)	(0.103)
x Initial R&D central.	-0.231	-0.227	-0.373+	-0.231	-0.229	-0.242
	(0.146)	(0.142)	(0.226)	(0.143)	(0.143)	(0.153)
R&D centralization of firm in initial period	-0.045	-0.031	-0.036	-0.032	-0.032	-0.036
	(0.031)	(0.027)	(0.052)	(0.028)	(0.027)	(0.030)
Expected ethnic share for RDHQ	0.115	0.048	-0.050	0.046	0.035	0.093
	(0.096)	(0.091)	(0.177)	(0.091)	(0.091)	(0.093)
Observations	12222	12222	12222	12222	12222	12222
R-Squared value	0.211	0.175	0.172	0.176	0.175	0.194
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes
Linear combination:						
30% centralization	0.008	0.018	-0.057	0.020	0.020	0.027
60% centralization	0.130+++	0.116+++	0.133+++	0.117+++	0.116+++	0.139+++
90% centralization	0.251+++	0.214+++	0.324+++	0.214+++	0.212+++	0.252+++

Notes: See Table 3.

Appendix Table 6c: Interaction estimations with sample splits

	Locations with >20 patent-inventors	Locations with >50 patent-inventors	Firms with single RDHQ that is >50% of patents	Firms with above median assignee filing decentr.	Firms with median and below assignee filing decentr.	Utilizing PatentsView assignee sample
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	0.014	-0.062	-0.218	-0.056	-0.075	0.116
	(0.084)	(0.094)	(0.146)	(0.130)	(0.094)	(0.077)
x Initial R&D central.	0.201+	0.274++	0.503++	0.279	0.360+++	0.121
	(0.117)	(0.134)	(0.198)	(0.185)	(0.135)	(0.098)
Expected ethnic share for satellite location	0.874+++	1.079+++	0.523++	0.681+++	0.796+++	0.799+++
	(0.154)	(0.161)	(0.214)	(0.151)	(0.127)	(0.074)
x Initial R&D central.	-0.147	-0.302	0.067	-0.117	-0.283	-0.173
	(0.234)	(0.239)	(0.281)	(0.210)	(0.193)	(0.108)
R&D centralization of firm in initial period	-0.015	-0.021	-0.130++	-0.018	-0.039	0.007
	(0.048)	(0.057)	(0.051)	(0.046)	(0.035)	(0.026)
Expected ethnic share for RDHQ	0.149	0.293++	0.184	0.168	0.122	0.051
	(0.099)	(0.136)	(0.160)	(0.122)	(0.153)	(0.067)
Observations	4715	2411	5084	6051	6171	17019
R-Squared value	0.370	0.504	0.214	0.212	0.226	0.198
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes
Linear combination:						
30% centralization	0.074	0.021	-0.067	0.027	0.032	0.152+++
60% centralization	0.134+++	0.103+++	0.084+	0.111++	0.140+++	0.188+++
90% centralization	0.194+++	0.185+++	0.235+++	0.195+++	0.249+++	0.224+++

Notes: See Table 3.

Appendix Table 6d: Interaction estimations with sample splits, continued

	Splitting the sample of technologies by the national rate of patenting growth among all inventors		Splitting the sample of technologies by the national rate of patenting growth among ethnic inventors		Splitting the sample of firms by their total span of distinct technologies across satellites	
	Median and above	Below median	Median and above	Below median	Median and above	Below median
	(1)	(2)	(3)	(4)	(5)	(6)
	DV is the ethnic share of inventors for the focal satellite R&D location					
Ethnic share of RDHQ	-0.077	-0.096	-0.065	-0.117	-0.074	-0.171
	(0.104)	(0.117)	(0.104)	(0.117)	(0.108)	(0.133)
x Initial R&D central.	0.345++	0.355++	0.327++	0.393++	0.262++	0.519++
	(0.143)	(0.168)	(0.144)	(0.170)	(0.128)	(0.209)
Expected ethnic share for satellite location	0.783+++	0.715+++	0.797+++	0.701+++	0.867+++	0.615+++
	(0.125)	(0.141)	(0.125)	(0.144)	(0.134)	(0.143)
x Initial R&D central.	-0.287	-0.176	-0.284	-0.175	-0.317	-0.153
	(0.183)	(0.200)	(0.182)	(0.205)	(0.203)	(0.205)
R&D centralization of firm in initial period	-0.030	-0.025	-0.029	-0.029	0.005	-0.078
	(0.047)	(0.033)	(0.046)	(0.034)	(0.040)	(0.050)
Expected ethnic share for RDHQ	0.063	0.196	0.061	0.210+	0.151	0.167
	(0.125)	(0.127)	(0.125)	(0.126)	(0.147)	(0.126)
Observations	6199	6023	6170	6052	7415	4807
R-Squared value	0.223	0.190	0.224	0.187	0.229	0.210
MSA and Tech-Pd FE	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes
Linear combination:						
30% centralization	0.027	0.011	0.330	0.000	0.005	-0.150
60% centralization	0.131+++	0.117+++	0.132+++	0.118+++	0.084+	0.140+++
90% centralization	0.234+++	0.224+++	0.230+++	0.236+++	0.162+++	0.296+++

Notes: See Table 3.

Appendix Table 7: Descriptive statistics for legacy sample

	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)
Count of distinct satellite locations	2898			
Count of satellite locations - time periods	4142			
… Inventor-patent pairs	99.968	479.392	5	10821
… Ethnic inventor share	0.178	0.211	0.000	1.000
… Expected ethnic inventor share	0.225	0.110	0.000	0.872
… Female inventor share	0.087	0.151	0.000	1.000
… More than 1500 mi. from RDHQ	0.232	0.422	0.000	1.000
… Different NBER subcategory vs RDHQ	0.595	0.491	0.000	1.000
… Rate of collaborating with RDHQ	0.323	0.325	0.000	1.000
… Rate of intra-firm career mobility to RDHQ	0.117	0.211	0.000	1.000
Count of firms	126			
… Ethnic inventor share of RDHQ	0.237	0.100	0.000	0.669
… Expected ethnic inventor share of RDHQ	0.212	0.046	0.104	0.402
… Minimum RDHQ share of firm patenting	0.533	0.141	0.316	0.961
… Share in ICT fields	0.343	0.475	0.000	1.000
R&D centralization in initial period	0.654	0.162	0.316	0.983

Notes: See Table 1. Legacy sample requires companies be active in patenting during 1975-1984 and have satellite locations after 1995.

Appendix Table 8: Table 7 with Period FE

	Baseline regression	Locations within 500 miles or direct plane flight of RDHQ	Locations NOT within 500 miles or direct plane flight of RDHQ	Locations with same tech focus as RDHQ	Locations with different tech focus than RDHQ	Separating firms by initial R&D centralization (30%, 60%]	(60%, 75%]	(75%, 100%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DV is the ethnic share of inventors for the focal satellite R&D location							
Ethnic share of RDHQ during 1975-1984	0.210+++	0.240++	0.179+++	0.200++	0.206+++	0.001	0.190+	0.523++
	(0.065)	(0.095)	(0.067)	(0.097)	(0.077)	(0.082)	(0.101)	(0.239)
Expected ethnic share for satellite location	0.667+++	0.751+++	0.584+++	0.579+++	0.727+++	0.747+++	0.598+++	0.713+++
	(0.059)	(0.085)	(0.099)	(0.104)	(0.074)	(0.072)	(0.106)	(0.115)
Expected ethnic share for RDHQ	0.040	0.073	0.038	-0.011	0.073	0.225	0.110	-0.247
	(0.114)	(0.151)	(0.144)	(0.158)	(0.185)	(0.260)	(0.204)	(0.192)
Observations	4142	1491	2651	1676	2466	1522	1576	1044
R-Squared value	0.220	0.296	0.235	0.259	0.270	0.321	0.310	0.266
MSA and Pd FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear difference of column pairs (2-3, 4-5)		0.061		-0.007				
		(0.102)		(0.119)				

Notes: See Table 7.

Appendix Table 9a: Patent quality of satellite facilities compared to RDHQ

	(0,1) Patent has a novel technology combination	Originality metric of patent	Log value of stock market reaction	(0,1) patent among top 10% stock market reaction for its grant year
	(1)	(2)	(3)	(4)
Satellite facility	-0.009+++	0.001	-0.036+++	-0.007+++
	(0.002)	(0.002)	(0.006)	(0.001)
Observations	1098596	1091899	1028348	1028348
R-Squared value	0.277	0.109	0.650	0.375
Firm-Period FE	Yes	Yes	Yes	Yes
MSA-Tech-Pd FE	Yes	Yes	Yes	Yes
Team Size FE	Yes	Yes	Yes	Yes

Notes: Estimations consider the traits of patents developed during a period at a firm and location. The estimations include an indicator variable for a satellite location and firm-period fixed effects, thus comparing satellite locations against their parent RDHQ. Estimations control for MSA-Tech-Period fixed effects, are unweighted, and cluster standard errors by firm-MSA-period.

Appendix Table 9b: Patent citations of satellite facilities compared to RDHQ

	Log value of patent citations	(0,1) patent among top 10% citations for its grant year	Log value of internal-to-firm patent citations	Log value of external-to-firm patent citations
	(1)	(2)	(3)	(4)
Satellite facility	-0.014	-0.001	-0.057+++	0.008
	(0.009)	(0.002)	(0.009)	(0.010)
Observations	1140679	1140679	1140679	1140679
R-Squared value	0.294	0.140	0.192	0.314
Firm-Period FE	Yes	Yes	Yes	Yes
MSA-Tech-Pd FE	Yes	Yes	Yes	Yes
Team Size FE	Yes	Yes	Yes	Yes

Notes: See Appendix Table 9a.

Appendix Table 10a: Patent quality of satellite facilities compared to each other

	(0,1) Patent has a novel technology combination	Originality metric of patent	Log value of stock market reaction	(0,1) patent among top 10% stock market reaction for its grant year
	(1)	(2)	(3)	(4)
(0,1) Satellite linked to RDHQ via collaboration/mobility	0.005+++	-0.002	0.020+++	0.003++
	(0.002)	(0.002)	(0.005)	(0.001)
Log count of patents	0.004+++	-0.001++	0.009+++	0.001
	(0.001)	(0.001)	(0.002)	(0.001)
Observations	458965	458291	409411	409411
R-Squared value	0.284	0.121	0.645	0.375
Firm-Period FE	Yes	Yes	Yes	Yes
MSA-Tech-Pd FE	Yes	Yes	Yes	Yes
Team Size FE	Yes	Yes	Yes	Yes

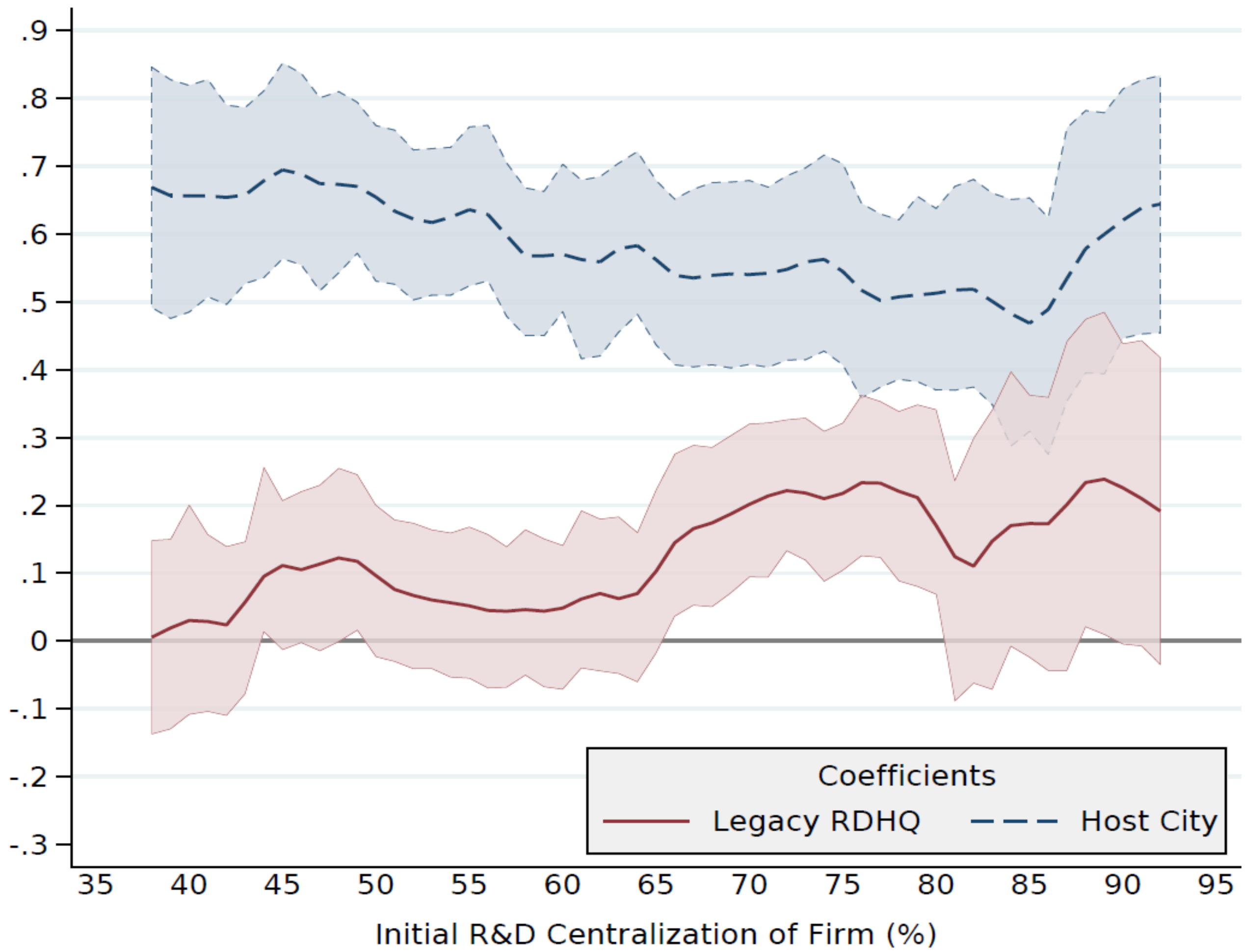
Notes: See Table 4a. Patent sample is restricted to satellite locations only.

Appendix Table 10b: Patent quality of satellite facilities compared to each other

	Log value of patent citations	(0,1) patent among top 10% citations for its grant year	Log value of internal-to-firm patent citations	Log value of external-to-firm patent citations
	(1)	(2)	(3)	(4)
(0,1) Satellite linked to RDHQ via collaboration/mobility	0.036+++	0.003+	0.099+++	0.001
	(0.009)	(0.002)	(0.007)	(0.009)
Log count of patents	0.001	-0.001++	0.027+++	-0.011+++
	(0.003)	(0.001)	(0.003)	(0.003)
Observations	477978	477978	477978	477978
R-Squared value	0.313	0.158	0.182	0.338
Firm-Period FE	Yes	Yes	Yes	Yes
MSA-Tech-Pd FE	Yes	Yes	Yes	Yes
Team Size FE	Yes	Yes	Yes	Yes

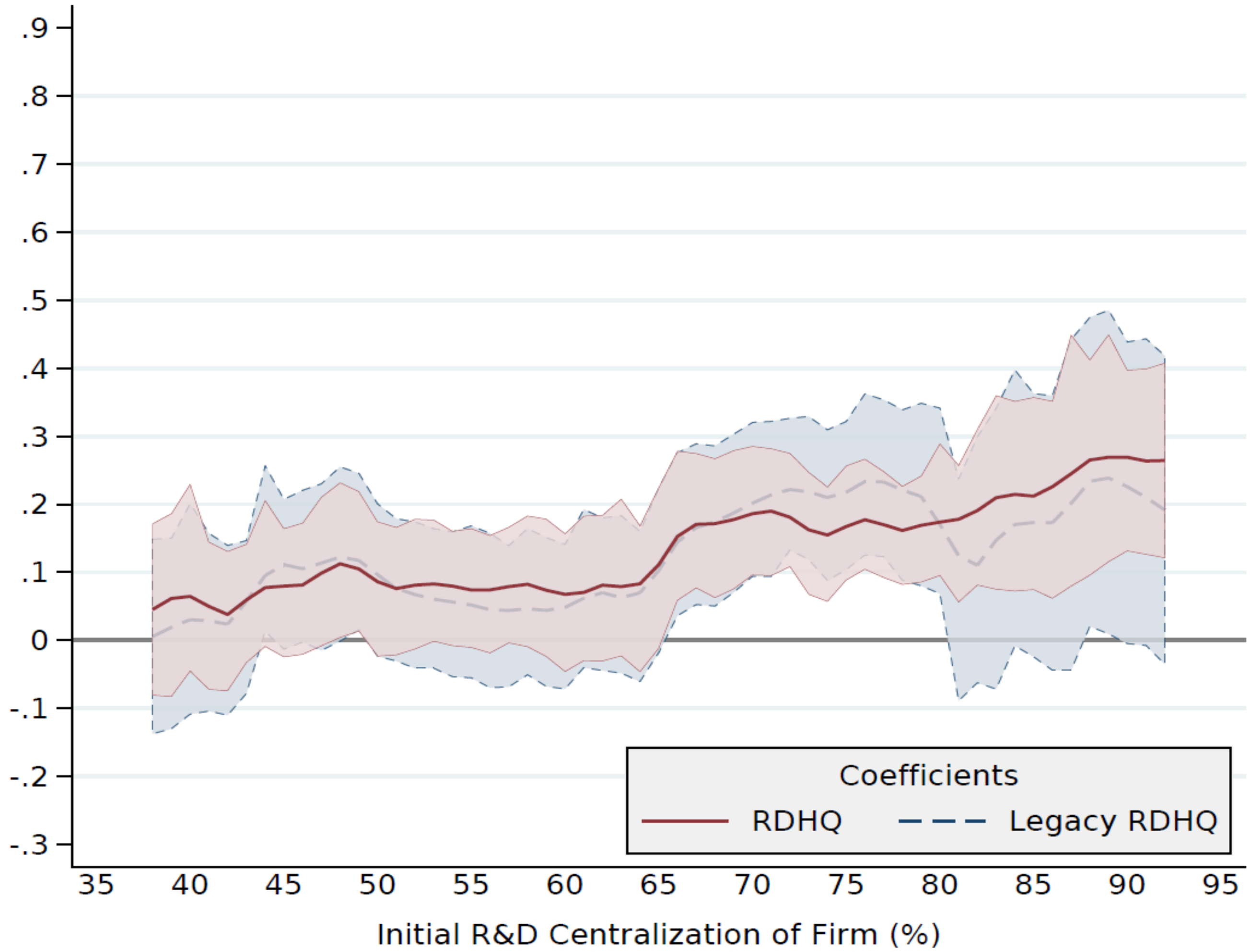
Notes: See Appendix Table 10a.

Appendix Figure 1: Explanatory power of legacy RDHQ workforce



Notes: See Figure 2 and Table 7.

Appendix Figure 2: Explanatory power of contemporary and legacy RDHQ workforce



Notes: See Figure 2 and Table 7.