The Role of the Firm in Nurturing Human Capital: Evidence from 200 Years of U.S. Patents

Xiaoding Liu
xliu@mays.tamu.edu
Texas A&M University
College Station, TX 77843

March 2025

ABSTRACT

In this paper, I study the increasing role of firms in developing human capital by adopting a long-run perspective on the entire evolution of the U.S. innovation landscape since the First Patent Act of 1790. I create a novel dataset of all U.S. patents granted since 1790 by processing over four million digitized patent images using state-of-the-art optical character recognition (OCR) conversion and machine learning based textual analysis. Examining over 200 years of U.S. patenting history reveals a striking fact: the share of total U.S. patents produced by corporations has increased from less than 1% in the early 1800s to over 90% today. Motivated by this remarkable shift in the locus of innovation production, I merge my proprietary patent dataset with the confidential full-count U.S. Census data from 1850 to 1940 to track the personal and family characteristics, as well as the employment, collaboration, migration, and productivity of the universe of U.S. inventors over a long horizon. Using this merged patent-inventor dataset, I present three novel findings regarding the role of firms in nurturing human capital. First, firms act as a critical source of initial employment for underrepresented R&D-focused workers (e.g., ethnic minorities and immigrants). Second, compared to self-employed independent inventors, corporations promote greater collaboration, mentorship, and knowledge sharing among their employee inventors. Third, corporate employee inventors, especially when working in teams, develop more impactful innovations compared to the patents developed by more homogenous networks of independent inventors. Based on these findings, I propose a new theory explaining the growing importance of firms in driving technological innovation that is centered on their role as 'knowledge agglomerators', which enables firms to maintain a lasting competitive advantage by facilitating the aggregation of dispersed knowledge, especially from underutilized skilled workers, and by establishing a cumulative knowledge base that employees can efficiently share and build upon over time.

Keywords: Innovation, Theory of the Firm, Corporate Culture, Organization of Innovation, Diversity, Inventor Teams, Labor Productivity, Entrepreneurship

JEL Classification: J24; L2; M14; M54; O31; O34

I. INTRODUCTION

As identified by Hayek (1945), the key underlying problem facing any economy is that the set of useful knowledge and innovative ideas necessary to drive economic growth are dispersed across a wide range of distinct, self-interested individuals. Therefore, one of the key goals of society is to establish an economic system that efficiently elicits and combines the disparate knowledge of separate individuals to develop socially beneficial and commercially viable economic outputs. The promotion of effective information sharing and knowledge recombination is especially important in driving technological innovation (Jones, 2009), where innovation is widely considered to be one of the most critical determinants of long-term economic prosperity (Schumpeter, 1942; Romer, 1990).

Given this background motivation, my paper investigates one of the most fundamental questions in economics, namely why do firms exist in such an economic system (Coase, 1937; Hart, 1988)? I propose that this question cannot be answered without understanding the critical role of firms in mitigating information coordination frictions and facilitating the human capital accumulation process. Specifically, I examine how firms can help to create an effective ecosystem that both: (a) attracts a broad range of talented individuals with diverse raw knowledge inputs and (b) fosters a workplace environment that promotes knowledge sharing between such individuals.

To study this broader research question, I focus on the most important input into the innovation production process, namely the individual inventors seeking to create new technologies, and analyze the entire evolution of the U.S. innovation landscape from the first official patent granted after the enactment of the First Patent Act of 1790. In particular, I create a novel dataset covering over 200 years of U.S. patenting history which identifies and links all the inventor/s *and* the assignee/s of each U.S. patent granted since 1790. To do so, I process over four million digitized patent images using state-of-the-art optical character recognition (OCR) conversion technology and machine learning based textual analysis algorithms. I then merge this proprietary patent dataset with the confidential full-count U.S. Census data from 1850 to 1940, enabling me to become the first researcher to construct a comprehensive database that tracks the personal characteristics, family attributes, and professional experiences of over one million individual U.S.-based inventors. This unique Census-patent matched database also allows me to track the migration patterns, collaboration choices, and productivity profiles of inventors during a critical multi-century period of U.S. economic and social change, which encompass, for example, multiple phases of the Industrial Revolution, two World Wars, the Great Depression, and unprecedented changes in U.S. civil rights.

My long-run analysis initially reveals three critical facts about the evolution of the U.S. innovation landscape over the last 200 years. The first critical fact is that there has been a *marked shift in patenting activity from self-employed 'independent inventors' to firms*. In particular, the share of total U.S. patents produced by employee inventors working within firms has grown from less than 1% in the early 1800's to over 90% today (see Figure 1).

In conjunction with the rising dominance of firms in the innovation production process, there have been two other important concurrent changes in the structure of the U.S. innovation ecosystem. The second critical fact is that the rising importance of teamwork in spurring new inventions is demonstrated by the *rapid growth in the size and diversity of inventor teams*, where the percentage of patents produced by teams of two or more inventors has increased from less than 5% in the early 1800s to over 70% today (see Figure 2). The third critical fact is that there has been an *extraordinary increase in the breadth and diversity of the inventor workforce*, where the percentage of inventors from a "minority" background (i.e., is a female or is from an ethnic minority group) has risen from less than 0.5% in the early 1800s to over 20% today (see Figure 3; see also Kerr, 2010; Foley and Kerr, 2013; and Fitzgerald and Liu, 2024).

Motivated by these dramatic changes, my study adopts a long-run perspective so that I can better understand how the U.S. innovation ecosystem evolved into its current form. This is in contrast to prior studies that rely on more limited datasets that only cover recent time periods when these key characteristics of the modern innovation eco-system (i.e., the dominance of firms, the high prevalence of teamwork, and the remarkable breadth and diversity of the inventor workforce) were already firmly established.[1]

More specifically, my study formally investigates the causal inter-relationship between firms, teamwork, and the aggregation and synthesization of inventor knowledge. If innovation is viewed primarily as a knowledge recombination problem (e.g., Fleming, 2001), then there are two potential methods by which firms could gain a sustainable competitive advantage compared to self-employed independent inventors. The first potential source of competitive advantage that firms could develop in the innovation process is by accumulating a broader and more diverse initial knowledge base, for example by accessing new and possibly underutilized sources of skilled human capital. This could in turn allow firms to build a more comprehensive range of raw knowledge inputs that can be used to generate more innovative ideas. A second potential source of competitive advantage for firms would be to create mechanisms to more efficiently combine raw information inputs into new commercial technologies. For example, corporate innovation outcomes may improve relative to self-employed inventors if firms were able to facilitate more efficient knowledge sharing among its employees and/or more effectively synthesize the differing perspectives of employee inventors.

Therefore, in the context of understanding the extraordinary growth in the importance of firms in the innovation production process, my empirical analysis seeks to answer the following research questions. First, are individuals from underrepresented backgrounds relatively more likely to be employed as inventors inside of firms or work as self-employed 'independent inventors'? Since these underrepresented individuals are different from mainstream inventors and thus more likely to

[1] This is because most prior (patent-based) innovation research studies rely upon publicly available USPTO digitized datasets whose coverage only starts in 1976.

possess different sets of knowledge and perspectives, this analysis enables me to gain insight into the broader question of whether firms are more effective at collecting and aggregating more disparate technical knowledge sources. Second, do firms promote more teamwork, mentorship, and knowledge sharing among their employees relative to groups of comparable self-employed inventors? This speaks to the broader question of whether firms are more efficient at integrating diverse knowledge inputs to create commercially useful technologies. Finally, depending on the answers to the first two questions, what are the implications for the innovation output of firms?

Using my unique patent–inventor–employer linked dataset and multi-century research setting, I present three novel findings about the long-term evolution of the U.S. innovation ecosystem.

First, *firms appear to act as a critical source of initial employment for underrepresented R&D-focused workers* (e.g., ethnic minorities and immigrants) outside of the traditional, geographically concentrated clusters of existing migrant communities. For example, I find that an inventor from a "minority" background is more likely to initially work as an employee of a company than white male inventors who were born in the United States if they do not have pre-existing access to a local network of ethnically similar entrepreneurs. Relatedly, I also show that the composition of inventors employed in corporate R&D units tends to be much more diverse in terms of ethnicity, immigrant status, and geography compared to contemporaneous cohorts of independent inventors. These findings are consistent with the anecdotal historical evidence that firms tended to implement relatively more systematic and open hiring practices compared to the network of independent inventors that relied more heavily on personal or family connections.

Second, I demonstrate that *corporations promote much greater collaboration, mentorship, and knowledge sharing among their employees*. In particular, I find that inventors are much more likely to collaborate with at least one other employee inventor working inside the same firm compared to otherwise similar individuals working as self-employed inventors. Furthermore, I show that firms especially promoted more teamwork between employee inventors from differing geographic, ethnic, gender, and/or immigrant backgrounds. Specifically, I estimate that corporate R&D teams are far more likely to comprise mixed gender, mixed ethnicity, and/or mixed immigrant-nonimmigrant teams compared to teams of independent inventors that tend to have a much more homogenous composition. As such, firms appear to significantly help inventors from underrepresented backgrounds to combine their knowledge, skills, and experiences with a more diverse array of inventors to pursue joint research endeavors, rather than be restricted to collaborations within their relatively homogenous existing networks of inventors with similar personal characteristics.

Third, I show that *corporate employee inventors, especially when working in teams, produce (on average) more impactful innovations* compared to the patents developed by more homogenous networks of self-employed independent inventors. In other words, I establish that the broader and more inclusive knowledge integration process within firms leads to significant differences in both the productivity and the types of innovation pursued by skilled individuals operating inside versus outside the boundaries of the firm. Specifically, I find that firms helped to produce a disproportionately high share of breakthrough inventions relative to the patents developed by independent inventors outside of firms. Importantly, this effect is largely driven by team-based patents produced within firms rather than solo employee inventor patents.

Crucially, existing economic theories describing the long-run development of innovation ecosystems cannot seem to fully explain this observed combination of empirical trends. In particular, I employ a multi-faceted identification strategy that can account for the potential selection effects and omitted variables concerns stemming from these existing economic theories. For example, through the inclusion of County × Time fixed effects, I can control for the influence of time-varying local economic conditions on the employment choices and productivity of inventors, which includes major economic shocks such as the Great Depression (see e.g., Babina, Bernstein, and Mezzanotti, 2023). Furthermore, through the inclusion of Technology class × Time fixed effects, I can control for the effect of changes in the nature and complexity of technological development over time, even within the same technology field (see e.g., Teece, 1988; Lamoreaux and Sokoloff, 2005).

As a result, I propose a new theory based on the pivotal role of firms as "knowledge agglomerators" that can jointly explain these seminal trends in the development of the modern innovation ecosystem. Under this theory, there are two distinct but interrelated components of firms that uniquely position them to act as efficient agglomerators of information and ideas, namely the principle of value maximization and the concept of corporate culture. First, as separate legal entities, firms are usually expected to endure indefinitely, where the well-accepted corporate governance principle is that the overall goal of the firm and its managers is to maximize shareholder value (Berk and DeMarzo, 2023). The second critical, yet often overlooked, aspect of the firm is that it also has its own distinct social identity, as embodied in its corporate culture. According to O'Reilly and Chatman (1996), corporate culture is "a system of shared values, beliefs, and norms that define appropriate attitudes and behaviors for organizational members." The reason that a firm's culture is so critical to a firm's operations is that corporate culture acts like an "invisible hand" that coordinates and guides employees on how to make decisions on matters that cannot be regulated or contracted on ex ante (e.g., Li, Mai, Shen, and Yan, 2021; Graham, Grennan, Harvey, and Rajgopal, 2022).

In my historical context, there are several reasons to believe that these unique characteristics of firms did have, and will continue to have, a profound impact on the development of technological innovation. First, consistent with the strong evidence of homophily in many small group settings,[2] it appears that "who you know" matters as much as, if not more than, "what you know" when attempting to establish oneself in traditional independent inventor networks (e.g., Bell, Chetty,

[2] For example, see Ishii and Xuan (2014), Gompers, Mukharlyamov, and Xuan (2016), and Calder-Wang, Gompers, and Huang (2023) in the context of directors/executives, venture capitalists, and firm founding teams, respectively.

Jaravel, Petkova, and Van Reenen, 2019). In contrast, the firm's overarching goal of value maximization can help focus a firm's initial recruiting decisions on an applicant's fundamental characteristics, such as their knowledge, skills, and experiences, rather than other extraneous aspects of an applicant's background, such as their ethnicity and immigrant status. Second, based on the concept of "identity economics" (Akerloff and Kranton, 2000), an important implication of corporate culture is that employees who share a vested interest in the firm's future prosperity will be more willing than otherwise similar independent inventors to make personal sacrifices for the betterment of others, even in the absence of monetary incentives (Gorton and Zentefis, 2024). For example, a senior R&D scientist may be more likely to exert costly effort to mentor and train a junior researcher when they are both working at the same company. Relatedly, because of their shared values and affinity for advancing their employer's long-term mission, employee inventors may be more willing to share their unique knowledge and overcome differences in their respective backgrounds and perspectives in order to collaborate with one another.

Therefore, I argue that the combined empirical evidence is consistent with my theory that the ability of firms to act as efficient knowledge agglomerators is directly responsible for firms steadily gaining a competitive advantage in developing commercially impactful technological innovations. Specifically, in their role as knowledge agglomerators, firms are able to perform two important functions. First, firms can effectively aggregate dispersed knowledge by implementing more institutionalized and equitable processes for selecting a broader set of skilled workers, including those from more marginalized communities. Second, through their corporate culture, firms can facilitate knowledge sharing between employee inventors by promoting teamwork and mentorship among a more diverse set of employee inventors, while also creating a cumulative knowledge repository that makes each new invention easier to develop. This is particularly important in the case of skilled individuals from minority or disadvantaged backgrounds that otherwise may not have had the necessary networks or resources to become a self-employed 'independent inventor' in the first place and thus may not have had the opportunity to learn from and collaborate with a broader range of inventors on new R&D projects.

An important implication of my findings is that firms appear to be a critical mechanism for allowing skilled individuals from underrepresented minority groups to overcome potential barriers to realizing their potential. As a result, I offer a new explanation for the central role of the firm in the modern innovation ecosystem, namely its superior ability to access underutilized skilled labor and to foster greater collaboration and knowledge sharing among a more diverse group of inventors.

My paper's findings contribute to several strands of literature. First, my results contribute to our understanding of the determinants and organization of innovative activities. To date, the prior literature has primarily focused on the effect of various economic, regulatory, industry, or firm-level factors on aggregate innovation production[3] and its consequent effect on the comparative advantage of for-profit corporations in managing innovative activities.[4] In contrast, my paper focuses on the actual unit of innovation production, namely the individual inventors tasked with developing new technologies, and the long-term evolution of their relationship with changing institutional structures surrounding the innovation production process. I provide novel historical evidence demonstrating the dual role of the firm in nurturing human capital talent and fostering an environment that encourages their employees to combine a more diverse set of intra-firm knowledge resources, thus helping to enhance overall organizational productivity. Furthermore, rather than being forced to take

[3] See, for example, determinants of innovation output such as economic and financing cycles (e.g., Nanda and Rhodes-Kropf, 2013; Huber, 2018; Babina et al., 2023), immigration patterns (e.g., Kerr and Lincoln, 2010; Moser, Voena, and Waldinger, 2014; Bernstein, Diamond, Jiranaphawiboon, McQuade, and Pousada, 2022), government investments in R&D (e.g., Gross and Sampat, 2023), taxation (e.g., Akcigit, Grigsby, Nicholas, Stantcheva, 2022), intellectual property laws (e.g., Moser, 2005; Mezzanotti, 2021), competition (e.g., Aghion, Bloom, Blundell, Griffith, and Howitt, 2005; Acemoglu, Akcigit, Alp, Bloom, and Kerr, 2018), mergers and acquisitions (e.g., Bena and Li, 2014; Seru, 2014; Frésard, Hoberg, and Phillips, 2020), and firm ownership structure (e.g., Aghion, Van Reenen, and Zingales, 2013).
[4] For a more general discussion, see for example: Teece (1988); Hughes (2004); Lamoreaux and Sokoloff (2005); and Arora, Belenzon, Kosenko, Suh, and Yafeh (2022).

as given the current features of the modern innovation eco-system, I can instead understand how the U.S. innovation eco-system evolved into its current form. For example, my study's long-run perspective allows me to document the full transition of underrepresented population groups, namely ethnic minority and immigrant researchers, as they gradually entered and established themselves in the skilled labor force, and to assess their impact on long-term economic development.

Second, my findings shed light on the debate surrounding the potential costs and benefits of increased workplace diversity, especially in R&D intensive environments. To date, most studies focus on the effect of more aggregate measures of diversity on aggregate performance, for example the effect of Board diversity or firm-wide diversity on firm innovation output.[5] In contrast, I examine how inventor- and team-level diversity in professional experiences and inherit traits can influence R&D productivity in different organizational settings. I show that the greater ability of corporations to produce more impactful patents compared to independent inventors in recent time periods can be at least partially attributed to firms being able to effectively hire and train individuals from a more diverse range of backgrounds, as well as firms being better able to promote knowledge sharing and collaboration between a more diverse set of employee inventors. My results illustrate more broadly the potential social and economic benefits of incorporating underrepresented inventor cohorts into a broader array of technological problem-solving activities.

Finally, my study relates to the broader question of the role of firms in promoting economic and social development. For instance, there remains ongoing public debate about the positive and negative effects, both direct and indirect, of actions undertaken by for-profit corporations and the degree of regulation necessary to protect the interests of various firm stakeholders. As it specifically

[5] With respect to the effect of Board diversity on corporate innovation, see e.g., An, Chen, Wu, and Zhang (2021) and Griffin, Li, and Xu (2021). With respect to the impact of firm-wide employee diversity on firm-level innovation, see e.g., Ostergaard, Timmermans, and Kristinsson (2011); Doran, Gelber, and Isen (2022).

relates to technological innovation, the question of whether it is optimal to organize R&D activities inside of firms is both theoretically ambiguous (e.g., Aghion and Tirole, 1994; Gromb and Scharfstein, 2002; Frésard et al., 2020) and empirically unclear (e.g., Lamoreaux, Sokoloff, and Sutthiphisal, 2009; Nicholas, 2010; Landes, Mokyr, and Baumol, 2012).

However, in order to address such questions, one must first understand the inner workings of corporations as well as delineate the exact nature of the relationship between firms and individuals. For example, relatively little is known about what explains whether aspiring inventors are more likely to work inside firms or pursue entrepreneurship (Babina et al., 2023) as well as how individual inventors interact with one another in different organizational settings. My study provides surprising evidence for the notion that firms are an important force in overcoming systematic barriers to entry into inventorship and alleviating the negative consequences of societal discrimination. By providing a critical source of initial employment and training for minority inventors, firms can generate long-lasting positive effects on the human capital of current employees and their subsequent offspring (see generally, Bell et al., 2019; Babina and Howell, 2024).

The rest of the paper is organized as follows. Section II outlines the data and variables used in my analysis. Section III studies the sorting of skilled researchers, particularly those from underrepresented backgrounds, into firm vs. non-firm environments. Section IV explores the dynamics of teamwork, mentorship, and knowledge sharing inside vs. outside of corporations. Section V examines the performance implications of these differences in the organization of innovative activities while Section VI concludes.

II. DATA AND VARIABLES

2.1 Historical patent data

To answer the research questions, I create my own historical patent database by extracting key inventor and assignee information from scanned patent images. I then match this database with the restricted full-count decennial U.S. census data from 1850 to 1940 to obtain additional inventor characteristics and track inventors over time. In the following subsections, I describe the creation of the historical patent database and the process of matching it with census data.

2.1.1 Digitizing historical patent records using Amazon OCR technology

I obtain nearly four million original PDF files of U.S. utility patents, spanning from Patent No. 1 to Patent No. 4,000,000, from the USPTO patent image database and the Google Patents database. These PDFs are scanned copies of the original patent documents, which were previously in paper form. U.S. Patent No. 1, issued to John Ruggles for a traction wheel for steam locomotives, was granted on July 13, 1836. Prior to the establishment of the U.S. Patent Office's numbering system in 1836, the office issued 9,957 unnumbered patents, the first of which, labeled “X” patents, was granted on July 31, 1790. However, most of these records were destroyed in a fire on December 15, 1836. As I do not have access to scanned PDF files of the X patents, they are not included in my main sample.

The quality of the scanned PDF files of patent records varies, with earlier patents typically having poorer quality. To extract key information, such as inventor names and locations, as well as assignee names and locations, I first convert these scanned PDFs into text files using Amazon AWS's Textract OCR service. AWS Textract is a comprehensive OCR solution that leverages machine learning models to extract text, forms, and tables from scanned documents. Patent PDFs present unique challenges due to their varying columnar structure, both within individual documents and across years. For example, the heading section is usually a single column, while the main text section is split into two columns. I have tested OCR conversion using Google Tesseract and several commercial services, but their output quality is significantly worse than AWS Textract, especially for more challenging scanned files.

2.1.2 Extracting inventor and assignee information using machine learning algorithms

Once the scanned patent documents are converted to raw text files, the next step is to extract key information, mainly from the header section of the document. There are several challenges associated with this information extraction process. First, identifying the header page is not simple because the header page can either be the first page of the patent document or follow pages of figures (whose number varies significantly across patents). Second, header formats have changed over time. For example, as shown in Figure 1A, the key words "United States Patent Office" is followed by inventor and assignee information, whereas these key words in Figure 1B is followed by invention name. These format variations make it more difficult to detect the start and the end of inventor and assignee information. Third, distinguishing inventor names from location names is not straightforward. The easier cases are when city names are preceded by "of". However, for a significant portion of the patent documents (especially in later years), inventor names and city names are separated by commas without unique identifying words. Fourth, when one or more inventors are deceased at the time of patent filing, the header section also includes administer/executor information. Since there is no standard format for such cases, it poses a challenge to distinguish inventors from administers.

Once the scanned patent documents are converted into raw text files, the next step is to extract key information, primarily from the header section. Several challenges arise in this extraction process. First, identifying the header page is not straightforward, as it could either be the first page of the patent document or follow pages of figures, whose number varies across patents. Second, header formats have evolved over time. For example, as shown in Figure 1A, the phrase "United States Patent Office" is followed by inventor and assignee information, whereas in Figure 1B, it is followed by the invention name. These format variations complicate the detection of the start and end of inventor and assignee information. Third, distinguishing inventor names from location names is challenging. In simpler cases, city names are preceded by the word "of" as in Figure 1A. However, in many patents, especially from later years, inventor names and city names are separated by commas without distinct identifying words as shown in Figure 1B. Finally, when one or more inventors are deceased at the time of patent filing, the header includes administrator or executor information. Since there is no standard format for such cases, it becomes difficult to distinguish between inventors and administrators.

Figure 1A. An example of the header section of a patent record.

Patented Mar. 17, 1925. **1,529,903**

UNITED STATES PATENT OFFICE.

JULIUS J. MOJONNIER AND HARLEY R. PHILLIPS, OF OAK PARK, AND HENRI B. WARREN, DECEASED, LATE OF OAK PARK, ILLINOIS; BY LEWIS W. WARREN, ADMINISTRATOR, OF OAK PARK, ILLINOIS, ASSIGNORS TO MOJONNIER BROS. CO., OF CHICAGO, ILLINOIS, A CORPORATION OF ILLINOIS.

CLUTCH.

Original application filed July 13, 1922, Serial No. 574,642. Divided and this application filed March 7, 1924. Serial No. 697,479.

Figure 1B. Another example of the header section of a patent record.

UNITED STATES PATENT OFFICE

2,480,166

RESISTOR FOR THERMOGAUGES

Karl Schwartzwalder, Holly, and Alexander S. Rulka and Robert W. Smith, Flint, Mich., assignors to General Motors Corporation, Detroit, Mich., a corporation of Delaware

Application January 8, 1945, Serial No. 571,812

5 Claims. (Cl. 201—76)

To address these issues, I implement the information extraction process in four steps. In the first step, I identify the header page of the patent document and extract the header section. To simplify subsequent steps, I also separate the inventor and assignee sections of the header. In the second step, I write extraction algorithms for more standard header formats, such as cases where inventor information follows the phrase "United States Patent Office" and city names are preceded by "of." I create different algorithms based on the number of occurrences of the keywords "and" and "of," accounting for variations in the number of inventors and locations. For example, when there are three inventors, the locations could vary in different ways: each inventor may come from a different location, two inventors might share a location, or all three could be from the same location. In the third step, for less standard formats, I use OpenAI machine learning algorithms to detect inventor and location names, verifying their accuracy manually. In the final step, a team of research assistants manually extracts information from the more difficult cases, such as patents with more than two inventors or deceased inventors.

After processing the inventor part of the header section, I move on to extracting assignee names and location information from the assignee section. I use a comprehensive list of keywords (e.g., "company," "corp," and "inc") to differentiate individual assignees from firm assignees. Once the firm assignee names are cleaned and standardized, I assign a unique firm ID to each firm-name-location observation.

2.2 Formation of inventor and assignee panel dataset

To obtain inventor characteristics beyond name and location information and be able to assign unique inventors IDs across patents and time, I match U.S. patents granted from 1836 to 1949 to the restricted full-count decennial U.S. census data from 1850 to 1940, which represent the complete set of Census records available to the public in which the respondents' names are disclosed. These census datasets contain more than 500 million individual records with detailed individual and household information.

I match the key inventor information extracted from the process described in Section 2.1.2 to the U.S. census records using a three-step process as follows:

In the first step, I clean inventor and city names to correct small OCR errors and spelling mistakes. I then split inventor names into first, middle, and last names. Similarly, I clean and split names in the census records in the same way. Since census records have consistent county names but not city names, I obtain inventor county information in two ways. First, I extract county details directly from the patent document, which is usually available for earlier patents where the first paragraph often includes both city and county names. Second, for patents lacking county information in the document, I use NHGIS county-level boundary files to identify the corresponding counties for inventor cities.

In the second step, I match each inventor-city-year observation to the census dataset that is closest to the patent grant year. For example, patents granted from 1935 to 1949 are matched to the 1940 census and patents granted from 1925 to 1934 are matched to the 1930 census. In all cases, I match an inventor only to individuals in the census who would be 17 years old or older at the time the patent is granted to the inventor. The matching is performed in four rounds. In round 1, matching is performed based on exact state, exact county, exact first name, exact last name, and exact middle name initial. Only individuals with the same non-missing middle name initial in both the historical patent database and the census database are matched. In round 2, matching is performed based on exact state, exact county, exact first name, exact last name, and non-conflicting middle name. Non-conflicting middle names refer to cases where an individual without a middle name is matched with an individual with a middle name. In round 3, matching is performed based on exact state, exact county, the last name with the highest similarity score of at least 95 out of 100, the first name with the highest similarity score of at least 95 out of 100, and exact middle name initial. In round 4, matching is performed based on exact state, exact county, the last name with the highest similarity score of at least 95 out of 100, the first name with the highest similarity score of at least 95 out of 100, and a non-conflicting middle name. From these four rounds, I am able to uniquely match around

50% of the inventors. When I require exact state matching instead of both state and county matching (allowing inventors to move to a different county within the same state) and use occupation types to filter out unlikely matches, the matching rate increases to around 56% on average.

In the third step, I compare the matched inventors from each census dataset to identify whether the same inventor is matched to multiple census datasets. To do so, I match across census records from different decades based on last middle, middle name, first name, gender, birth year, place of birth, race, parents' names, and parents' birthplace (if available). By matching across census datasets, I can track inventors over time, even if they moved to a different county or state. Once the matching is complete, I generate a unique inventor ID for each inventor.

2.3 Variable construction

In this section, I describe the independent and dependent variables used in my empirical analysis. All patent-related information comes from the historical patent database that I created (see Section 2.1), while all inventor characteristics including age, gender, race, place of birth, and immigrant status come from the restricted full-count decennial U.S. census data. Appendix A provides further details on the construction of each of these variables.

2.3.1 Inventor-level variables

I define several variables to capture key inventor characteristics. *Immigrant inventor* is an indicator that equals one if the inventor is a first-generation immigrant, and zero if the inventor is born in the United States. *Minority Inventor* is an indicator that equals one if an inventor belongs to one of the ethnic minority groups (i.e., Hispanic, Black, Asian or Pacific Islander, and American Indian), and equals zero if an inventor identifies as White. *Male* is an indicator that equals one for male inventors and zero for female inventors. *Age at Time of Patent* is the age of an inventor at the time the patent is granted. Following Jones (2009), I use age at first invention as a measure of the amount of knowledge and experience accumulated by the focal inventor to date.

2.3.2 Patent-level variables

I first define three variables that capture key characteristics of the granted patent. The first variable is *Company Assignee* or *Company Patent*, which is an indicator that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only. The second variable is *Team Patent*, which is an indicator that equals one if the patent has more than one inventor and zero if the patent has only a single inventor. The third variable is *Number of Inventors*, which is the number of inventors listed in the patent document.

In terms of measuring the quality of the focal patent output, I use two measures based on the patent text-based methodology of Kelly, Papanikolaou, Seru, and Taddy (2021). The first variable is *Patent Importance*, which is defined as the ratio of 10-year forward patent text similarity (i.e., the aggregate patent text similarity between a patent and all patents granted in the next ten years) to the 5-year backward patent text similarity (i.e., the aggregate patent text similarity between a patent and all patents granted in the prior five years). Therefore, this measure defines important patents as patents that are both novel compared to prior patents (i.e., low backward patent text similarity) and impactful to future patents (i.e., high forward patent text similarity). In other words, these patents represent a significant improvement over prior technology and lay the groundwork for future inventions. The second variable is *Breakthrough Patent*, which is a dummy variable that is equal to one if the patent falls in the top 10% of the unconditional distribution of the patent importance measure, and zero otherwise.

I also construct several empirical measures to capture the diversity of the focal R&D team. *Mixed Immigrant Team* captures team diversity in terms of immigrant status. This indicator equals one (and zero otherwise) if the team has at least one immigrant inventor and one non-immigrant inventor. *Mixed Minority Team* captures team diversity in terms of (ethnic) minority status. This indicator equals one (and zero otherwise) if the team has at least one minority inventor and one non-minority inventor. To capture the geographic diversity of the R&D team, *Number of Different States in a team* refers to the count of distinct U.S. states where the inventors on a team are located and *Number of Different Counties in a team* refers to the count of distinct U.S. counties where the inventors on a team are located.

2.4 Summary statistics

Table 1 presents key summary statistics from my historical patent-assignee linked dataset. In Panel A, I outline the demographic characteristics of the individual inventors in my sample which spans from 1845 to 1950. In general, patenting inventors tended to be White males, with approximately one-fifth of inventors being first-generation immigrants.

In Panel B of Table 1, I provide further details about inventors' mode of employment and their collaboration choices. I initially document that approximately one-third of inventors in my Census-matched sample were employed at R&D-focused firms. I also find that approximately 85% of patents during my Census-matched sample period were produced by individual inventors working alone, while the remainder were produced by teams of two or more inventors.

III. EVOLUTION OF EMPLOYMENT CHOICE BY U.S. BASED INVENTORS

In this section, I examine the evolution in the selection of employment mode by U.S. based inventors over time. Specifically, I seek to identify the key factors that affect how inventors from differing personal and professional backgrounds choose whether to operate as a self-employed 'independent inventor' or work as an employee of an R&D focused firm.

3.1 Baseline determinants of employment mode selection in realized inventor sample

In my first set of selection tests, I use the actual sample of inventors that received at least one patent during my sample period to understand their initial choice of whether to work as a corporate R&D employee or operate as an independent inventor.[6] Following the prior literature, I define an inventor's employer as the firm that is the patent assignee. To obtain an objective estimate of an inventor's entry into the skilled R&D labor force, I utilize the application date of that inventor's first (eventually granted) patent application.

3.1.1 Baseline specification

I use the following linear probability model regression specification to isolate the key factors that influence whether an inventor decides to commence their career by becoming an employee of a corporate R&D facility:

$$\begin{aligned}\mathbb{I}\{Company\ Patent_{i,t}\} \\ &= \alpha + \beta_1\ \%\ of\ same\ minority/immigrant\ inventor\ in\ county_i \\ &\quad + \beta_2\ Male_i + \beta_3\ Age\ at\ time\ of\ patent_{i,t} + State\ \times Time\ FEs \\ &\quad + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t} \qquad (1)\end{aligned}$$

$\mathbb{I}\{Company\ Patent_{i,t}\}$ is an indicator variable that is equal to one if a patent's assignee is a company (i.e., the corporate employer of the focal inventor), and zero otherwise. To account for the effect of time-varying local economic conditions as well as differences in the nature and complexity of developing inventions across technology fields, I include County × Time and Technology Class × Time fixed effects, respectively. I run the analysis using two separate subsamples, where the first subsample consists of first-generation immigrant inventors and the second subsample consists of

[6] I focus on the very first patent that an individual inventor receives in my sample because the choice of workplace is less likely to be affected by confounding factors such as prior experience and team-specific relationship capital. I examine the factors that influence an inventor to switch between being an independent inventor or a corporate R&D employee in later robustness tests.

inventors belonging to one of the ethnic minority groups (i.e., Hispanic, Black, Asian or Pacific Islander, and American Indian). Please refer to Section 2 for the definition of all independent variables included in this analysis.

3.1.2 Baseline empirical results

The results of my initial selection tests are reported in Table 2. I first find that minority inventors that have a limited local network of individuals from a similar minority background to tap into are much more likely to initially seek employment within firms. Importantly, however, minority inventors who grow up in counties with a high percentage of individuals from the same minority background as the focal individual are significantly more likely to commence their career as an independent inventor. I find similar patterns for first generation immigrant inventors. More broadly, this finding is indicative of the importance of social networks in the inventor team formation process.

IV. NATURE OF COLLABORATION INSIDE AND OUTSIDE THE FIRM

In this section, I investigate the impact that R&D organizational structure, namely working inside the firm versus working as an independent inventor, has on the level of teamwork among comparable groups of inventors as well as the influence of organizational structure on the (inherited) diversity of inventor teams.

4.1 Baseline relationship between organizational structure and teamwork

I use the following regression specifications to isolate the key factors that influence whether inventors decide to work individually or as part of a team on a new patentable invention:

$$\mathbb{I}\{Team\ Patent_{i,t}\} = \alpha + \beta_1\ Corporate\ assignee_i \qquad (2a)$$

$$+ County\ \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t}$$

$$Number\ of\ inventors = \alpha + \beta_1\ Corporate\ assignee_i \qquad (2b)$$

$$+ County\ \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t}$$

My first dependent variable is $\mathbb{I}\{Team\ Patent_{i,t}\}$ which is an indicator variable that is equal to one if a patent's assignee is a company (i.e., the corporate employer of the focal inventor), and zero otherwise. My second dependent variable is $Number\ of\ inventors$ which is a continuous variable equal to the number of inventors listed on the patent document. My main independent variable of interest is $Corporate\ assignee_i$ which is an indicator variable equal to one if the patent was developed by employees working inside a corporation, and zero otherwise.

To account for the effect of time-varying local economic conditions as well as differences in the nature and complexity of developing inventions across technology fields on the need for teamwork, I include County × Time and Technology Class × Time fixed effects, respectively. As such, my empirical tests involve comparing sets of inventors working in the same geographic region at the same time (so thus exposed to the same local economic shocks) and in the same technology class (so thus exposed to similar technological and competitive landscape), but where one set of inventors is working inside a firm while the other set of individuals is working as independent inventors. Please refer to Section 2 for the definition of all independent variables included in this analysis.

In Table 3, I report the results of my empirical tests related to the determinants of inventor teamwork. I find that, even for sets of inventors working in the same geographic region at the same time and in the same technology class, corporate inventors are much more likely to work together in teams. This result suggests that firms play an important role in promoting knowledge sharing among corporate R&D employees.

4.2 Baseline relationship between organizational structure and team diversity

I use the following regression specification to isolate the propensity for more diverse teams to form inside vs. outside of the firm:

$$\mathbb{I}\{Mixed\ minority\ team_{i,t}\} = \alpha + \beta_1\ Corporate\ assignee_i + County \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t} \qquad (3a)$$

$$Number\ of\ different\ states\ or\ counties = \alpha + \beta_1\ Corporate\ assignee_i + County\ \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t} \qquad (3b)$$

My first dependent variable is $\mathbb{I}\{Mixed\ minority\ team_{i,t}\}$ which is an indicator variable that is equal to one if the inventor team includes at least one minority inventor and at least one non-minority inventor, and zero otherwise. My second dependent variable is $Number\ of\ different\ States$ or $Number\ of\ different\ counties$ which are continuous variables that equal the number of distinct U.S. States or the number of distinct U.S. counties where the inventors on a team are located. My main independent variable of interest is $Corporate\ assignee_i$ which is an indicator variable equal to one if the patent was developed by employees working inside a corporation, and zero otherwise. I include County × Time fixed effects and Technology class × Time fixed effects in all regressions.

I report the results of my empirical tests related to the characteristics of inventor teamwork in Table 4. I find that inventors with a more diverse array of inherited characteristics and a more diverse geographic locations are significantly more likely to collaborate with one another inside of a firm than outside of a corporate R&D setting. My results imply that firms can help to overcome barriers to collaboration between skilled individuals, both in terms of homophily based on inherited traits as well as physical proximity.

To further explore what types of teams are formed by self-employed independent inventors, I use the following regression specification to isolate the propensity for more homogenous teams to form inside vs. outside of the firm:

$$\mathbb{I}\{All\ minority\ team_{i,t}\} = \alpha + \beta_1\ Independent\ assignee_i + County \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t} \quad (3c)$$

$$Same\ state\ or\ county = \alpha + \beta_1\ Independent\ assignee_i + County\ \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t} \quad (3d)$$

My first dependent variable is $\mathbb{I}\{All\ minority\ team_{i,t}\}$ which is an indicator variable that is equal to one if the inventor team is the consisting of all minority inventors, and zero otherwise. My second dependent variable is $Same\ state$ or $Same\ county$, which are indicators that equal one if all inventors on the team come from the same state or county. My main independent variable of interest is $Independent\ assignee_i$ which is an indicator variable equal to one if the patent was developed by individuals, and zero otherwise. I include County × Time fixed effects and Technology class × Time fixed effects in all regressions.

I report the results of my empirical tests related to the characteristics of inventor teamwork in Table 5. I find that independent inventors are more likely to form homogenous teams from the same geographic location, which is consistent with the notion that these inventors tend to tap into a limited local network of individuals from a similar minority background.

4.3 Baseline relationship between organizational structure and mentorship

To test whether firms may promote more mentorship of less experienced employee inventors by more experienced employee inventors relative to groups of comparable independent inventors, I run the following regression specification:

$$\mathbb{I}\{Age\ gap\ greater\ than\ 5\ or\ 10\ years_{i,t}\}$$
$$= \alpha + \beta_1\ Corporate\ assignee_i + County \times Time\ FEs + Tech\ class \times Time\ FEs + \varepsilon_{i,t} \quad (4)$$

My main dependent variables of interest are large age gap dummies, where age gap refers to the difference between the youngest inventor and the oldest inventor on the team. My main independent variable of interest is $Corporate\ assignee_i$ which is an indicator variable equal to one if the patent was developed by employees working inside a corporation, and zero otherwise. I include County × Time fixed effects and Technology class × Time fixed effects in all regressions.

I report the results of my empirical tests related to the level of mentorship provided by more experienced inventors to less experienced inventors inside vs. outside of firm boundaries in Table 6. I find that inventors working inside of a firm are much more likely to collaborate with and mentor more junior corporate inventor colleagues while more experienced independent inventors are significantly less likely to partner with and mentor less experienced self-employed inventors. This result suggests that firms must have mechanisms to encourage more senior employees to share their technical information and institutional knowledge with more junior employees.

V. IMPLICATIONS OF TEAMWORK ON R&D PRODUCTIVITY

In this section, I examine the innovation outcomes that result from inventors working inside vs. outside of the boundaries of the firm, including the effect of teamwork and inventor team diversity on subsequent team-level innovation output.

5.1 The impact of firms and teams on innovation outcomes

My first set of treatment effects tests examine how impactful is the patent output of inventors, delineated by whether they are working inside or outside of firms and whether they are working individually or as part of a team. Specifically, we run the following regression:

$$Patent\ importance = \alpha + \beta_1\ Corporate\ assignee_i\ + \beta_2\ Team\ patent_i \quad (5)$$
$$+ \beta_3\ Corporate\ assignee_i \times Team\ patent_i$$
$$+ County\ \times Time\ FEs + Tech\ class\ \times Time\ FEs + \varepsilon_{i,t}$$

The dependent variable $Patent\ importance$ is equal to the ratio of 10-year forward patent text similarity (i.e., the aggregate patent text similarity between the patent and all patents granted in the next ten years) to the 5-year backward patent text similarity (i.e., the aggregate patent text similarity between the patent and all patents granted in the prior five years), following the methodology in Kelly, Papanikolaou, Seru, and Taddy (2021). Alternatively, I also use a *Breakthrough patent* indicator as a dependent variable, which equals one (zero otherwise) if the patent falls in the top 10% of the unconditional distribution of the patent importance measure. I include County × Year fixed effects and Technology class × Time fixed effects in all regressions.

In Table 7, I report the results of this empirical analysis. I first find that inventors working inside of firms in general produced more important patents compared to traditional networks of independent inventors. More crucially, however, I show in column (2) that patents produced by teams of corporate inventors were a particularly important source of competitive advantage for firms relative to the patents produced by independent inventors. This implies that it was not simply just access to greater financial or physical resources that led to firms producing more impactful patents relative to more financially constrained independent inventors. Instead, my results suggest that the promotion of greater teamwork and knowledge sharing within corporate R&D environments can at least partially explain the ability of corporations to eventually become the dominant contributor to aggregate innovation production in the modern knowledge economy.

VI. POTENTIAL THEORIES EXPLAINING EMPIRICAL RESULTS

In this section, I explore the relative merits of various theoretical frameworks that may be able to explain the significant changes observed in the U.S. innovation landscape over the past 200 years.

6.1 Existing theories

While various theories have been advanced to provide an explanation for some of the seminal trends that I document in the evolution of the modern innovation landscape, it is often difficult for these existing theories to *jointly* reconcile the existence and magnitude of these key trends.

6.1.1 Superior financing capacity of firms

With respect to the shift in patenting activity from independent inventors to firms, some have argued that the increase in firms' comparative advantage in innovation can be explained by the superior ability of firms to raise large pools of capital and finance costly R&D investments over longer time horizons (Hughes, 2004; Lamoreaux and Sokoloff, 2005). This is especially important during severe economic downturns such as the Great Depression (Babina et al., 2023).

However, if this is the primary explanation for all the various key historical trends that I document, then there are several inconsistencies in the empirical data that need to be reconciled. As an example, why didn't individual inventors simply join these better financially resourced firms and then continue to work on predominantly solo-authored R&D projects? In addition, why was the shift in the locus of patenting activity from independent inventors to firms already underway prior to the Great Depression and why didn't the patenting activity of independent inventors recover after the Great Depression, once angel financing market conditions had sufficiently improved?

6.1.2 Fundamental changes in the complexity of patentable technologies

Other studies have proposed that increases in the underlying complexity of developing new technologies can account for both the shift in patenting activity from independent inventors to firms as well as the general rise in teamwork among inventors. In particular, papers such as Jones (2009) assert that the rise in the number and size of inventor teams can be justified as a means for successive generations of individual inventors to more efficiently reduce their "knowledge burden" (i.e., their initial learning and training cycle in a new technical area). Given the increase in the specialized knowledge resources and physical capital needed to advance upon existing technologies, it is argued that firms developed a more enduring competitive advantage in managing more complex R&D projects relative to independent inventors (e.g., Lamoreaux and Sokoloff, 2005). However, this theory is not without its own challenges. For example, in an innovation-focused environment that is heavily reliant on tacit knowledge (Fitzgerald and Liu, 2024), what is the exact mechanism or organizational feature that enables firms to systematically foster more teamwork among employees relative to groups of independent inventors? In addition, this theory is conspicuously silent on the question of diversity in hiring decisions and inventor team formation choices, especially as it pertains to the significant differences observed between independent inventors and firms.

Nevertheless, to account for these potential alternative explanations, I include several layers of fixed effects and explicit control variables. With respect to the influence of relative financial constraints, my most granular tests include County × Time fixed effects. These fixed effects thus ensure that we compare the contemporaneous employment choices and productivity of inventors that are exposed to the same local economic conditions at the same time. With respect to the impact of changes in the complexity of developing new inventions across different technology classes, my most granular tests include Technology Class × Time fixed effects. These fixed effects thus ensure that we compare the decision-making and R&D outcomes of inventors who are working in the same technology field at the same time.

6.2 An alternative theory based on the role of firms as 'knowledge agglomerators'

Distinct from these existing theories, I propose a new theory based on the critical role of firms as "knowledge agglomerators" that can jointly rationalize the three seminal trends in the modern innovation ecosystem. Under this theory, there are two distinct but mutually reinforcing principles, namely the drive for shareholder value maximization and the development of a positive corporate culture, that are both necessary for a successfully functioning corporate environment.

With respect to the former principle, it is widely accepted that the overall objective of the firm, its managers, and its employees is to maximize firm value for the benefit of shareholders who are the ultimate owners of the company (Berk and DeMarzo, 2023). A related feature of the corporate structure is that firms have a legal identity that is separate from its shareholders. Importantly, firms are typically expected to operate indefinitely and they possess many of the same rights and responsibilities as natural persons, including the ability to enter into contracts, sue and be sued, and buy assets and incur liabilities in their own name.

However, the overarching objective of value maximization alone is not sufficient to overcome all coordination issues faced by individual workers in their day-to-day activities. This is because it is impossible for the managers of a firm to monitor and dictate each decision that employees will make as part of their core responsibilities (e.g., Li et al., 2021). This is especially true in the case of innovation production where individual inventors are confronted by complex R&D problems where the optimal end output or even the optimal search process for discovering a potential solution is often highly uncertain (D'Acunto, Tate, and Yang, 2021).

Therefore, the second critical, yet often overlooked, aspect of a firm is that it will develop its own distinct social identity (otherwise known as its "corporate culture"). According to O'Reilly and Chatman (1996), corporate culture is "a system of shared values, beliefs, and norms that define appropriate attitudes and behaviors for organizational members." As stated in several surveys of firm executives, corporate culture is like an "invisible hand" that acts as a "coordination mechanism" to guide employee decision-making on issues that cannot be regulated or contracted on ex ante (e.g., Graham, Grennan, Harvey, and Rajgopal, 2022). The critical role of corporate culture in influencing the strategic vision and daily operations of the firm is indicated by the fact that corporate executives rank culture as the most important contributor to long-term firm value (Graham et al., 2022).

The central role of corporate culture in determining economic output is likely to be accentuated in very open-ended and dynamic settings such as innovation production. To successfully tackle these challenging R&D projects, individual inventors need to accumulate and maintain a sufficiently high level of human capital (Jones, 2009). This learning and training process not only includes the acquisition of raw technical knowledge (so called "hard skills") but also the development of personal qualities and traits that impact how R&D projects are undertaken, such as communication, critical thinking, and problem-solving abilities (so called "soft skills"). However, unlike physical resources, human capital is (by its very nature) not easily acquired or transferred between individuals, with this continual learning and training process occupying a significant proportion of the innovation lifecycle (Jones, 2009). In such cases, formal contracts and compensatory incentives alone are unlikely to be sufficient to promote the development of the tools necessary to consistently produce commercially successful innovations (Aghion and Tirole, 1994).[7]

Based on the novel viewpoint of firms as efficient "knowledge agglomerators", I propose that this theory leads to a set of testable predictions about the innovation process inside vs. outside of firm boundaries. Specifically, I use the historical development of the modern U.S. innovation system over a multi-century period as a natural laboratory for examining how the unique characteristics of firms can influence the selection, professional interactions, and productivity of individual inventors across different R&D organizational structures.

[7] For example, if a piece of physical equipment is needed for a production process, then a contract for the purchase and assignment of this item to the relevant end user can be executed in a relatively expeditious manner. In contrast, if a person needs to learn about and train in a new technology area, then this requisite skillset cannot be simply bought and transferred to the targeted individual. Instead, this human capital accumulation process usually requires inputs from several different, more experienced individuals and often takes many years to complete.

First, in many different small group settings, there is strong evidence of 'homophily', namely the tendency of individuals to prefer to work with others who share similar personal characteristics (Ewens, 2023). This phenomenon was particularly evident in the relatively tight knit community of existing independent inventors and angel investors, where almost all independent inventors were White males from relatively privileged socioeconomic backgrounds. This was likely to have been a particularly imposing barrier to entry for minority inventors, especially those from humbler origins that did not have pre-existing social connections to geographically proximate independent inventors. Furthermore, given that independent inventors would tend to focus on specific R&D projects over more condensed time periods (somewhat analogous to the modern-day independent contractors), relationships in this eco-system were likely to have been more transactional and short-term in nature. This in turn might limit the opportunity for aspiring inventors to engage with and learn from more experienced independent inventors.

To alleviate the constraining effect of such biases, it is possible that minority inventors sought paid employment in corporate R&D laboratories. Unlike independent inventors, value-maximising firms were likely to be less concerned about a job applicant's immutable characteristics (e.g., their inherited traits like ethnicity, gender, family's socioeconomic status etc.) and more focused on the applicant's fundamental attributes (e.g., raw knowledge, skills etc.) as well as their long-term potential to contribute to the organization's future growth. A more systematic and equitable approach to hiring would have been particularly beneficial for aspiring inventors from minority or disadvantaged backgrounds.[8] In addition, firms also probably possessed greater ability to adopt a longer-term perspective on developing the human capital of employee inventors, providing employees with relatively more freedom to cultivate and apply their innate skills and knowledge to

[8] To be clear, I do not claim that no 'homophily' or no other discriminatory biases existed inside of firms, only that such biases were probably less prevalent inside of firms as compared to independent inventor networks.

a broader range of technological problems. These combined factors lead to my first set of empirically testable predictions, namely that:

> Prediction 1: *Conditional on access to existing networks of inventors with similar personal backgrounds, underrepresented or disadvantaged inventors are more likely to initially work as an employee inside of a firm rather than become a self-employed entrepreneur.*

Indeed, as shown in Table 2, my empirical results are consistent with the important role of firms in identifying and incorporating underrepresented individuals into the innovation ecosystem.

Second, another implication of 'homophily' in R&D intensive environments is that individuals may avoid collaborating with others from different personal backgrounds because, all else being equal, it will increase the likelihood of interpersonal conflicts and miscommunications (e.g., Garlappi, Giammarino, and Lazrak, 2017). According to the model of Gorton and Zentefis (2024), this preference can be expressed as a 'disutility' cost that individuals incur if they work with someone with differing values, beliefs, and perspectives. However, a unique countervailing force inside of a firm is that individuals also have a set of shared interests and beliefs that stem from their employer's corporate culture, such that employees will experience positive utility from helping their colleagues to maintain a productive and culturally harmonious workplace environment (Gorton, Grennan, and Zentefis, 2022). As a result, consistent with the theory of 'identity economics' by Akerloff and Kranton (2000), employees will make decisions based not only on monetary incentives but also the corporate identity to which they are socially connected, often leading them to make choices that align with their perceived group identity, even if it means foregoing some personal gain.

In a R&D focused setting, the unifying role of corporate culture can have several impacts on the innovation process. For example, more experienced inventors may be more willing to voluntarily share their knowledge and to help mentor and train the next generation of aspiring inventors when they are both working inside of the same company (Ackigit, Caicedo, Miguelez, Stantcheva, and

Sterzi, 2024). In addition, a shared corporate culture may help with the communication and synthesisation of differing ideas and perspectives, especially across individuals from disparate personal backgrounds, because shared cultural values can cut across ethnic, gender, geographic, and other personal boundaries. Furthermore, the greater opportunity for interaction between corporate R&D scientists and 'frontline' workers may facilitate the faster identification of real-world commercial problems, the collection of data, and the more efficient formulation of practical technological solutions. In contrast, given the more limited scope and lifespan of projects undertaken by independent inventors, these self-employed "guns for hire" may only perform the tasks that are necessary for executing the current project and thus may have less interest in mentoring and training aspiring inventors.

Since both culture and invention share a reliance on tacit knowledge (Fitzgerald and Liu, 2024), it is plausible that differences in the cultural environment inside vs. outside of firms played a major role in shaping the modern innovation ecosystem. As such, this leads to my second set of theoretical predictions that we can evaluate against the available empirical data, namely that:

> <u>Prediction 2a</u>: *There is more teamwork among inventors working inside of the same firm than among groups of otherwise similar independent inventors.*
>
> <u>Prediction 2b</u>: *There are more diverse teams working inside of a firm that comprise both minority and non-minority inventors than among groups of otherwise similar sets of independent inventors.*
>
> <u>Prediction 2c</u>: *There is more mentorship of less experienced inventors by more experienced inventors when working inside of the same firm than among groups of otherwise similar independent inventors.*

The empirical evidence presented in Table 3 strongly supports the theory that firms promote much greater collaboration among employee inventors, while the results of Table 4 illustrate that firms do indeed encourage much greater teamwork and knowledge sharing among a more diverse set of inventors relative to groups of otherwise comparable independent inventors. Furthermore, the empirical evidence that I present in Table 6 strongly indicates that more experienced inventors are much more willing to collaborate with and mentor less experienced inventors if they are both working inside the same firm.

Third, by accessing and combining a more diverse set of knowledge and perspectives, it is plausible that the more integrated corporate R&D process could led firms to develop more rigorously tested and impactful innovations compared to relatively more isolated independent inventors. This leads to my final set of theoretical predictions, namely that:

<u>Prediction 3</u>: *Firm patents, especially those generated by inventor teams, will be more impactful than patents produced by independent inventors.*

Indeed, the evidence presented in Table 7 suggests that more efficiently aggregating and integrating a more diverse set of knowledge inputs is very beneficial for producing more impactful, radical innovations. As such, my theory of firms acting as more efficient 'knowledge agglomerators' has the ability to jointly explain many of the critical, long-run empirical trends that I document.

VII. CONCLUSION

In this paper, I study the increasing role of firms in developing inventor human capital and its transformative impact on technological innovation. I propose that the extraordinary rise in the importance of firms in the modern innovation ecosystem can be attributed to the pivotal role firms in the knowledge agglomeration and human capital accumulation process. Specifically, I argue that each corporation's goal of maximizing shareholder value and fostering a unifying corporate culture enables them to create a workplace environment that is conducive to collaborative innovation production. To test this theory, I create a novel inventor-assignee linked database that tracks the inherited traits, acquired experiences, employment decisions, collaboration choices, and overall productivity of the universe of U.S. inventors over a multi-century period.

Using this novel dataset and research setting, I first find that firms act as a critical source of initial employment for underrepresented R&D-focused workers. Furthermore, I demonstrate that firms serve as a more efficient mechanism to facilitate teamwork, mentorship, and knowledge sharing among a more diverse set of inventors. As a result, I find that firms are increasingly better positioned to develop more impactful innovations compared to the more homogenous networks of self-employed independent inventors. My study's findings thus have important implications for the optimal structure of commercial innovation activities as well as the role of the firm in promoting more long-term orientated and merit-based workplace environments.

REFERENCES

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and W. Kerr. 2018. Innovation, Reallocation, and Growth. *American Economic Review* 108(111): 3450–3491.

Ackigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., and V. Sterzi, 2024. Dancing with the Stars: Innovation through Interactions. Working Paper, NBER.

Akcigit, U., Grigsby, J., Nicholas, T., and S. Stantcheva. 2022. Taxation and Innovation in the 20th Century. *Quarterly Journal of Economics* 137(1): 329–385.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and P. Howitt. 2005. Competition and Innovation: An Inverted-U Relationship. *Quarterly Journal of Economics* 120(2): 701–728.

Aghion, P., Van Reenen, J., and L. Zingales. 2013. Innovation and Institutional Ownership. *American Economic Review* 103(1): 277–304.

Aghion, P., and J. Tirole. 1994. The Management of Innovation. *Quarterly Journal of Economics* 109(4): 1185–1209.

Akerloff, G., and R. Kranton. 2000. Economics and Identity. *Quarterly Journal of Economics* 115(3): 715–753.

Arora, A., Belenzon, S., Kosenko, K., Suh, J., and Y. Yafeh. 2022. The Rise of Scientific Research in Corporate America. Working Paper, NBER.

Babina, T., Bernstein, A., and F. Mezzanotti. 2023. Financial Disruptions and the Organization of Innovation: Evidence from the Great Depression. *Review of Financial Studies* 36(11): 4271–4317.

Babina, T., and S. Howell. 2024. Entrepreneurial Spillovers from Corporate R&D. *Journal of Labor Economics* 42(2): 469–509.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., and J. Van Reenen. 2019. Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *Quarterly Journal of Economics* 134(2): 647–713.

Bena, J., and K. Li. 2014. Corporate Innovations and Mergers and Acquisitions. *Journal of Finance* 69(5): 1923–1960.

Berk, J., and P. DeMarzo. 2023. *Corporate Finance*. Hoboken, NJ: Pearson.

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

Coase, R.H. 1937. The Nature of the Firm. *Economica* 4(16): 386–405.

D'Acunto, F., Tate, G., and L. Yang. 2021. Entrepreneurial Teams: Diversity of Skills and Early-Stage Growth. Working Paper, University of Maryland.

Ewens, M. 2023. Gender and Race in Entrepreneurial Finance. Chapter 6 in *Handbook of the Economics of Corporate Finance* 1: 239–296.

Foley, F., and W. Kerr. 2013. Ethnic Innovation and U.S. Multinational Firm Activity. *Management Science* 59(7): 1529–1544.

Fitzgerald, T., and X. Liu. 2024. Shared Culture and Technological Innovation: Evidence from Corporate R&D Teams. Working Paper, Texas A&M University.

Fleming, L. 2001. Recombinant Uncertainty in Technological Search. *Management Science* 47(1): 117–132.

Frésard, L., Hoberg, G., and G. Phillips. 2020. Innovation Activities and Integration through Vertical Acquisitions. *Review of Financial Studies* 33(7): 2937–2976.

Garlappi, L., Giammarino, R., and A. Lazrak. 2017. Ambiguity and the Corporation: Group Disagreement and Underinvestment. *Journal of Financial Economics* 125(3): 417–433.

Graham, J., Grennan, J., Harvey, C., and S. Rajgopal. 2022. Corporate Culture: Evidence from the Field. *Journal of Financial Economics* 146(2): 552–593.

Gorton, G., and A. Zentefis. 2024. Corporate Culture as a Theory of the Firm. *Economica* 91(364): 1391–1423.

Gromb, D., and D. Scharfstein. 2002. Entrepreneurship in Equilibrium. Working Paper, NBER.

Gross, D., and B. Sampat. 2023. America, Jump-Started: World War II R&D and the Takeoff of the U.S. Innovation System. *American Economic Review* 113(12): 3323–3356.

Hall, B., and J. Lerner. 2010. The Financing of R&D and Innovation. In *Handbook of the Economics of Innovation* 1: 609–639.

Hayek, F.A. 1945. The Use of Knowledge in Society. *American Economic Review* 35(4): 519–530.

Hart, O. 1988. Incomplete Contracts and the Theory of the Firm. *Journal of Law, Economics, and Organization* 4(1): 119–139.

Huber, K. 2018. Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties. *American Economic Review* 108: 868–898.

Hughes, T. 2004. *American Genesis: A Century of Invention and Technological Enthusiasm, 1870–1970*. Chicago, IL: University of Chicago Press.

Jones, B. 2009. The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation getting harder? *Review of Economic Studies* 76(1): 283–317.

Kelly, B., Papanikolaou, D., Seru, A., and M. Taddy. 2021. Measuring Technological Innovation over the Long Run. *American Economic Review Insights* 3(3): 303–320.

Kerr, W. 2010. The Agglomeration of US Ethnic Inventors. In Glaeser, E. (ed). *Economics of Agglomeration* 1: 237–276. University of Chicago Press, Chicago.

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

Lamoreaux, N., and K. Sokoloff. 2005. The Decline of the Independent Inventor: A Schumpterian Story? Working Paper, NBER.

Lamoreaux, N., Sokoloff, K., and D. Sutthiphisal. 2009. The Reorganization of Inventive Activity in the United States during the Early Twentieth Century. Working Paper, NBER.

Landes, D., Mokyr, J., and W. Baumol. 2012. *The Invention of Enterprise: Entrepreneurship from Ancient Mesopotamia to Modern Times*. Princeton, NJ: Princeton University Press.

Li, K., Mai, F., Shen, R., and X. Yan. 2021. Measuring Corporate Culture Using Machine Learning. *Review of Financial Studies* 34: 3265–3315.

Mezzanotti, F. 2021. Roadblock to Innovation: The Role of Patent Litigation in Corporate R&D. *Management Science* 67(12): 7291–7950.

Moser, P. 2005. How do Patent Laws Influence Innovation? Evidence from Nineteenth Century World Fairs. *American Economic Review* 95: 1214–1236.

Moser, P., Voena, A., and F. Waldinger. 2014. German Jewish Émigrés and US Invention. *American Economic Review* 104(10): 3222–3255.

Nanda, R., and T. Nicholas. 2014. Did Bank Distress Stifle Innovation during the Great Depression? *Journal of Financial Economics* 114: 273–292.

Nanda, R., and M. Rhodes-Kropf. 2013. Investment Cycles and Startup Innovation. *Journal of Financial Economics* 110(2): 403–418.

Nicholas, T. 2010. The Role of Independent Invention in US Technological Development, 1880–1930. *Journal of Economic History* 70: 57–82.

O'Reilly, C., and J. Chatman. 1996. Culture as Social Control: Corporations, Cults, and Commitment. In B.M. Staw and L.L. Cummings (eds), *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews* 18: 157–200.

Romer, P. 1990. Endogenous Technological Change. *Journal of Political Economy* 98: 71–102.

Schumpeter, J. 1942. *Capitalism, Socialism and Democracy*. New York: Harper.

Seru, A. 2014. Firm Boundary Matters: Evidence from Conglomerates and R&D activity. *Journal of Financial Economics* 111: 381–405.

Teece, D. 1988. Technological Change and the Nature of the Firm. Chapter 12 in *Technical Change and Economic Theory* 1: 219–308.

Figure 1 – The transition of R&D activities from independent inventors to corporations

The figure shows the percentage of total U.S. patents produced by firms vs. independent inventors from 1790 to 2018. The underlying data come from my proprietary dataset of U.S. patents created from original patent images.

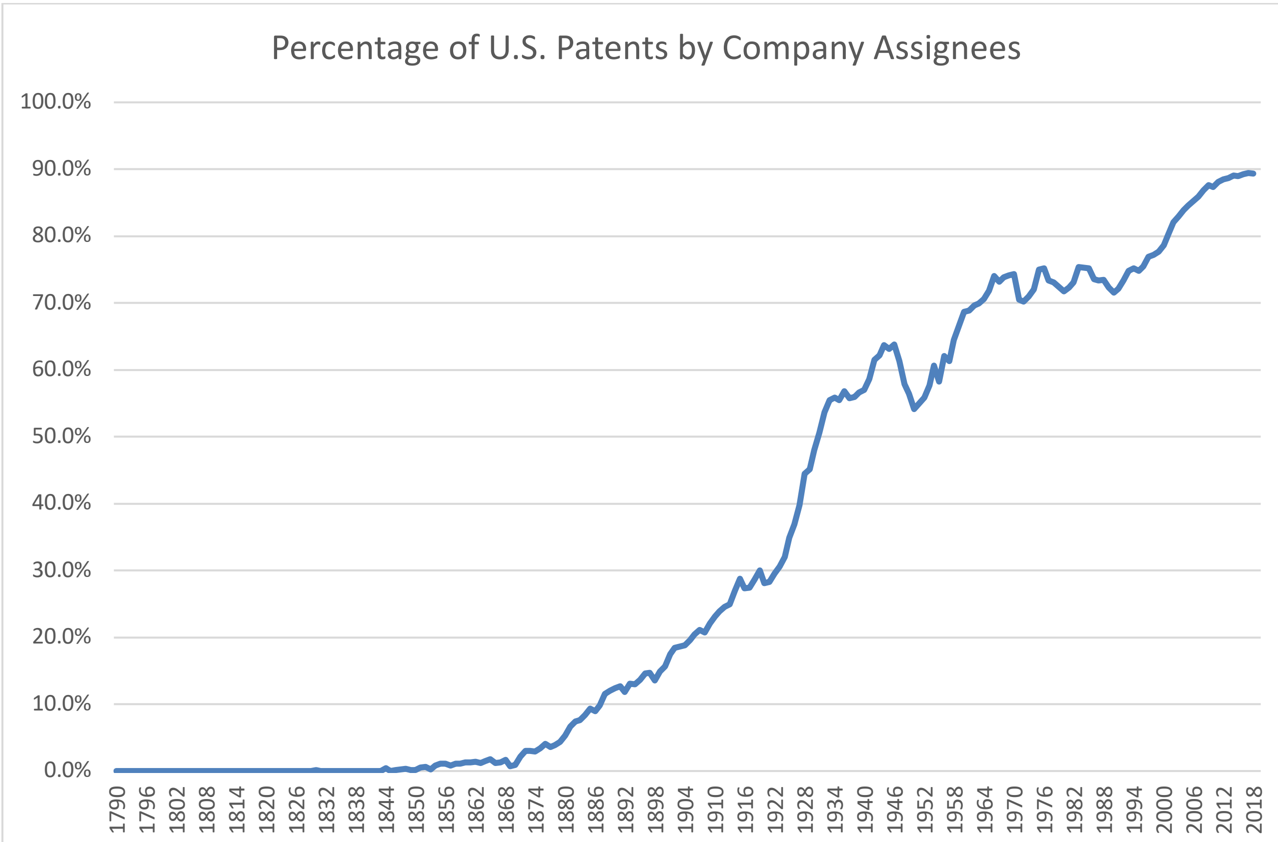

Figure 2 – The growth in the size of inventor teams

The figure shows the percentage of total U.S. patents produced by teams of more than one inventor and the average number of inventors per patent from 1790 to 2018. The underlying data come from my proprietary dataset of U.S. patents created from original patent images.

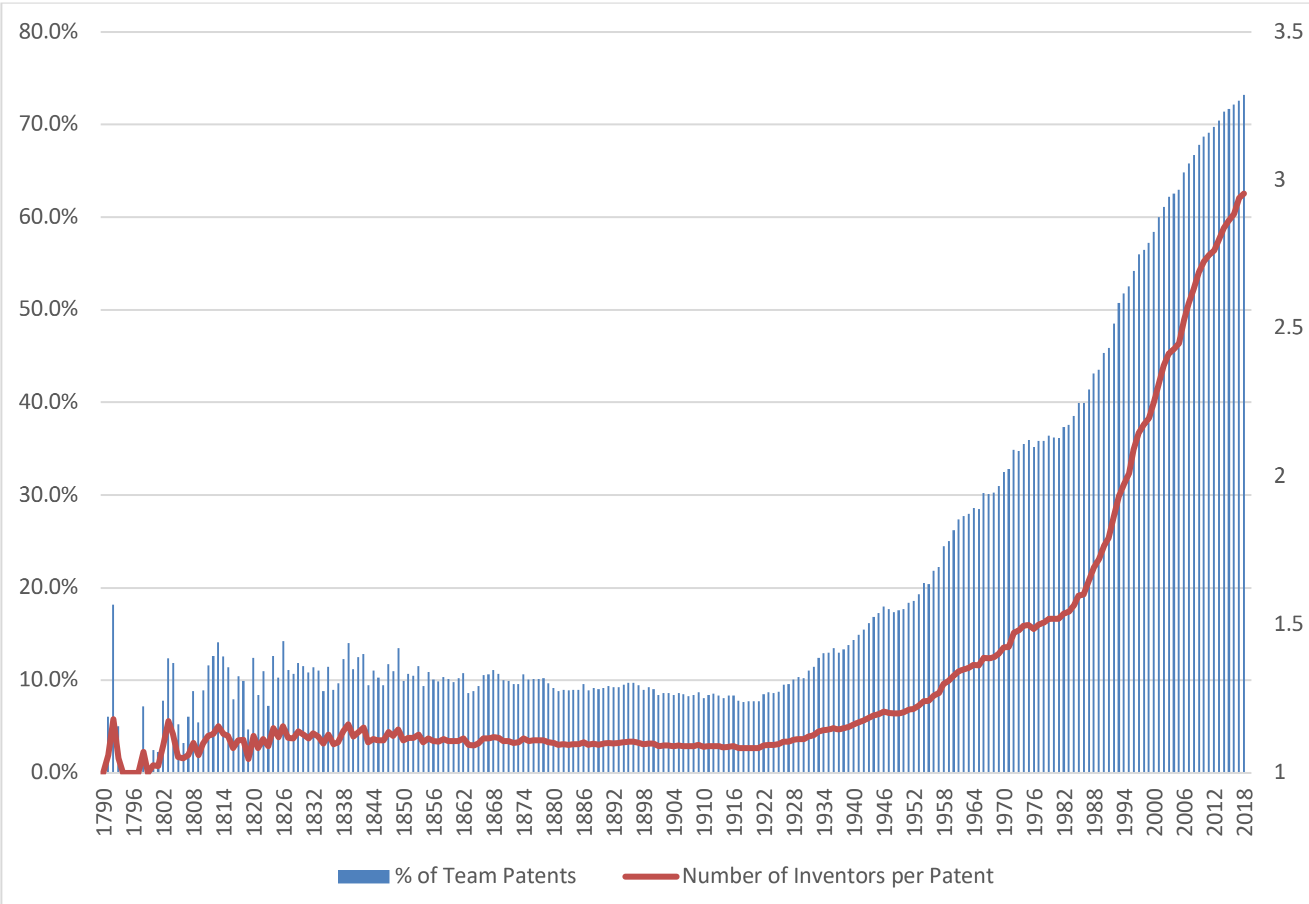

Figure 3 – The increasing diversity of the inventor population

The figure shows the percentage of total U.S. patents produced by female inventors or inventors belonging to one of the ethnic minority groups (i.e., Hispanic, Black, Asian or Pacific Islander, and American Indian) from 1845 to 1950. The underlying data come from my proprietary dataset of U.S. patents created from original patent images, which is then merged with the confidential full-count U.S. Census data from 1850 to 1940.

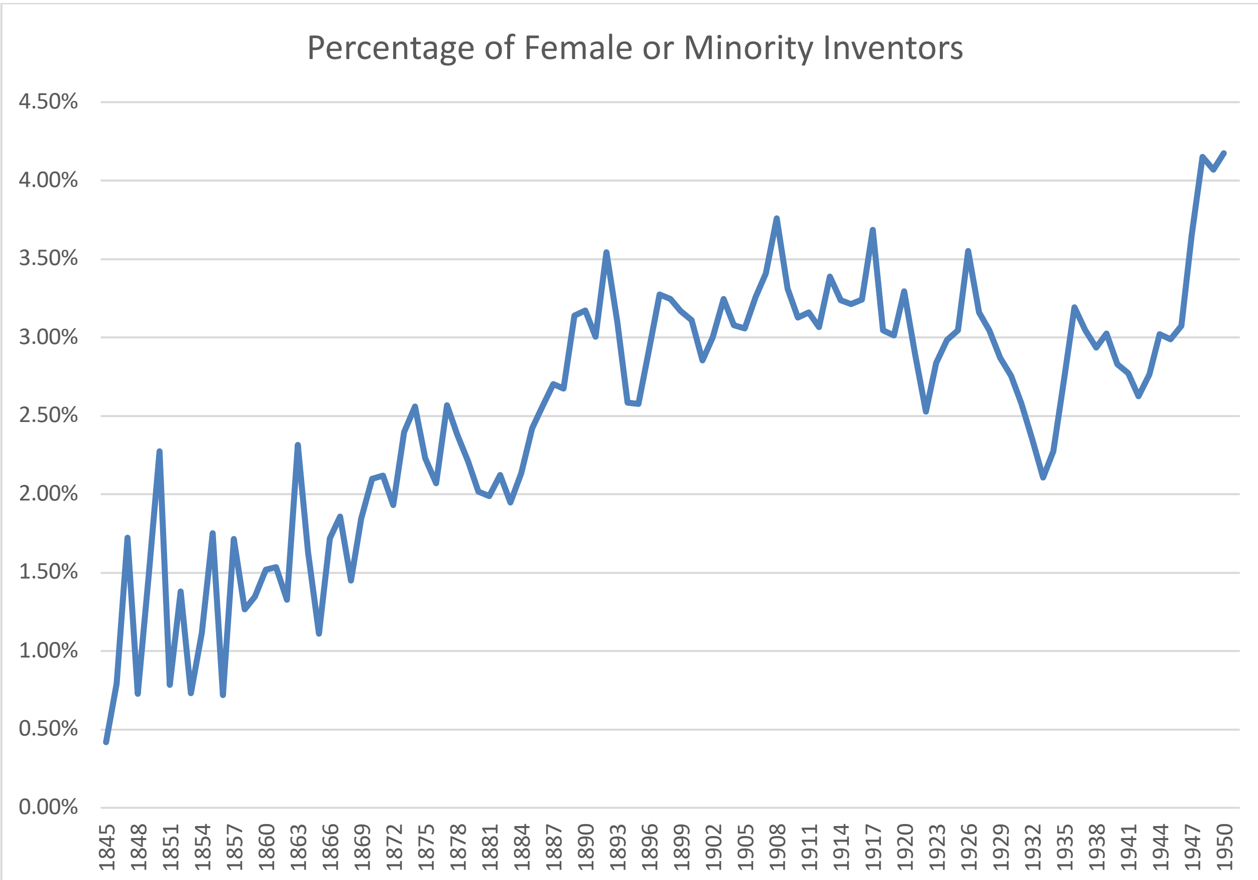

Table 1: Summary statistics

This table reports summary statistics for the matched census sample from 1845 to 1950. Panel A presents descriptive statistics at the inventor level. Panel B present descriptive statistics at the patent level. Panel C outlines descriptive statistics for inventor teams consisting of more than one inventor. Appendix A outlines the definition of all the variables listed.

Panel A: Individual inventor characteristics

	N	Mean	Median	Std. dev.
Male	1,372,959	0.987	1	0.112
Age at time of patent	1,372,959	43.228	42	12.726
Immigrant inventor	1,372,959	0.185	0	0.389
Minority inventor	1,372,959	0.017	0	0.129
% of immigrants from the same origin country as the inventor in the county	1,372,959	0.006	0	0.021
% of minority from the same minority group as the inventor in the county	1,372,959	0.003	0	0.031

Panel B: Patent characteristics

	N	Mean	Median	Std. dev.
Company assignee	1,284,371	0.341	0	0.474
Independent assignee	1,284,371	0.659	1	0.474
Team patent	1,284,371	0.142	0	0.349
Number of inventors	1,284,371	1.157	1	0.411
Patent importance	1,284,333	0.781	0.753	0.272
Breakthrough patent	1,284,333	0.069	0	0.254

Panel C: Team characteristics

	N	Mean	Median	Std. dev.
Mixed immigrant team	182,018	0.097	0	0.297
Mixed minority team	182,018	0.012	0	0.109
Mixed immigrant/minority team	182,018	0.107	0	0.309
Number of different states	182,018	1.107	1	0.326
Number of different counties	182,018	1.214	1	0.427
All immigrant team	182,018	0.123	0	0.421
All minority team	182,018	0.011	0	0.102
All immigrant/minority team	182,018	0.132	0	0.339
Same state team	182,018	0.889	1	0.315
Same county team	182,018	0.793	1	0.405

Table 2: Which inventors work in firms?

This table reports the regression results on how inventor characteristics relate to the decision to work in firms using patent-inventor level observations. The sample period is from 1845 to 1950. Company patent is an indicator that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only. Column (1) includes first-generation immigrant inventors in the sample and column (2) includes inventors belonging to one of the ethnic minority groups (i.e., Hispanic, Black, Asian or Pacific Islander, and American Indian) in the sample. % of immigrants from the same origin country as the inventor in the county is the percentage of individuals from the inventor's county that came from the same original native country as the immigrant inventor. % of minority from the same minority group as the inventor in the county is the percentage of individuals from the inventor's county belonging to the same ethnic minority group as the inventor. Male is an indicator that equals one for male inventors and zero for female inventors. Age at time of patent is the age of an inventor at the time of patent grant. State (county) refers to the inventor's location at the time of patent grant. Tech class refers to the patent's CPC technology class. *t*-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the inventor level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Dependent variable	Company patent	Company patent
	(1)	(2)
% of immigrants from the same origin country as the inventor in the county	-0.467***	
	(-9.33)	
% of minority from the same minority group as the inventor in the county		-0.074***
		(-2.68)
Male	0.175***	0.109***
	(19.25)	(6.70)
Age at time of patent	-0.0003*	-0.001
	(-1.84)	(-1.64)
State × Time FE	Yes	Yes
Tech class × Time FE	Yes	Yes
Number of Obs.	253,838	23,080
Adjusted R^2	0.243	0.325
Sample	First Generation Immigrant Inventors	Minority Inventors

Table 3: Are teams more likely to form inside or outside of firms?

This table reports the regression results on the relation between company assignee and team patenting using patent-level observations. The sample period is either the full sample from 1836 to 2018 or the census matched sample from 1845 to 1950. Panel A presents results on collaboration choices by organization type and Panel B presents results on team size by organization type. Team patent is an indicator that equals one if the patent has more than one inventor and zero if the patent has only a single inventor. Number of inventors is the number of inventors listed in the patent document. Company assignee is an indicator that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only. State (county) refers to the inventor's location at the time of patent grant and is based on the first inventor's location for patents with multiple inventors. Tech class refers to the patent's CPC technology class. *t*-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the county level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: Collaboration Choices by Organization Type

Dependent variable	Team patent	Team patent	Team patent	Team patent
	(1)	(2)	(3)	(4)
Company assignee	0.230***	0.209***	0.049***	0.050***
	(26.79)	(23.15)	(20.72)	(16.81)
State × Time FE	Yes		Yes	
County × Time FE		Yes		Yes
Tech class × Time FE	Yes	Yes	Yes	Yes
Number of Obs.	6,110,783	6,110,783	1,284,371	1,284,371
Adjusted R^2	0.274	0.290	0.027	0.037
Sample	Full Sample (1836-2018)		Census Matched Sample (1845-1950)	

Panel B: Team Size by Organization Type

Dependent variable	Number of inventors	Number of inventors	Number of inventors	Number of inventors
	(1)	(2)	(3)	(4)
Company assignee	0.527***	0.469***	0.060***	0.060***
	(22.01)	(19.25)	(21.01)	(23.45)
State × Time FE	Yes		Yes	
County × Time FE		Yes		Yes
Tech class × Time FE	Yes	Yes	Yes	Yes
Number of Obs.	6,110,783	6,110,783	1,284,371	1,284,371
Adjusted R^2	0.254	0.272	0.029	0.039
Sample	Full Sample (1836-2018)		Census Matched Sample (1845-1950)	

Table 4: Do firms promote more diverse teams?

This table reports the regression results on how patent assignee type relates to team characteristics using only patent-level observations with teams of more than one inventor. The sample period is from 1845 to 1950. Mixed immigrant team is an indicator that equals one (zero otherwise) if the team has at least one immigrant inventor and one non-immigrant inventor. Mixed minority team is an indicator that equals one (zero otherwise) if the team has at least one minority inventor and one non-minority inventor, where minority inventors refer to inventors who belong to one of the ethnic minority groups (i.e., Hispanic, Black, Asian or Pacific Islander, and American Indian). Mixed immigrant/minority team is an indicator that equals one (zero otherwise) if the team is classified as a mixed immigrant team or a mixed minority team. Number of different states in a team refers to the count of distinct U.S. states where the inventors on a team are located. Number of different counties in a team refers to the count of distinct U.S. counties where the inventors on a team are located. Company assignee is an indicator that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only. County refers to the inventor's location at the time of patent grant and is based on the first inventor's location for patents with multiple inventors. Tech class refers to the patent's CPC technology class. *t*-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the county level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Dependent variable	Mixed immigrant team	Mixed minority team	Mixed immigrant/ minority team	Number of different states in a team	Number of different counties in a team
	(1)	(2)	(3)	(4)	(5)
Company assignee	0.014***	0.003***	0.016***	0.051***	0.082***
	(4.97)	(3.81)	(5.45)	(4.09)	(5.51)
County × Time FE	Yes	Yes	Yes	Yes	Yes
Tech class × Time FE	Yes	Yes	Yes	Yes	Yes
Number of Obs.	182,018	182,018	182,018	182,018	182,018
Adjusted R^2	0.012	0.066	0.015	0.078	0.170
Sample	Census Matched Sample (1845-1950)				

Table 5: Do independent investors have more homogenous teams?

This table reports the regression results on how patent assignee type relates to team characteristics using only patent-level observations with teams of more than one inventor. The sample period is from 1845 to 1950. Mixed immigrant team is an indicator that equals one (zero otherwise) if the team has at least one immigrant inventor and one non-immigrant inventor. All immigrant team is an indicator that equals one (zero otherwise) if the team has all immigrant inventors. All minority team is an indicator that equals one (zero otherwise) if the team has all minority inventors. All immigrant/minority team is an indicator that equals one (zero otherwise) if the team is classified as an all-minority team or an all-immigrant team. Same state team is an indicator that equals one (zero otherwise) if all inventors come from the same state. Same county team is an indicator that equals one (zero otherwise) if all inventors come from the same county. Independent assignee is an indicator that equals one if a patent's assignee is an individual and zero otherwise. County refers to the inventor's location at the time of patent grant and is based on the first inventor's location for patents with multiple inventors. Tech class refers to the patent's CPC technology class. *t*-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the county level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Dependent variable	All immigrant team	All minority team	All immigrant/ minority team	Same State team	Same county team
	(1)	(2)	(3)	(4)	(5)
Independent assignee	0.037***	0.002	0.038***	0.042***	0.076***
	(5.33)	(1.40)	(5.35)	(4.88)	(5.61)
County × Time FE	Yes	Yes	Yes	Yes	Yes
Tech class × Time FE	Yes	Yes	Yes	Yes	Yes
Number of Obs.	182,018	182,018	182,018	182,018	182,018
Adjusted R^2	0.058	0.100	0.060	0.089	0.171
Sample	Census Matched Sample (1845-1950)				

Table 6: Are firms more likely to promote mentorship between inventors?

This table reports the regression results on how patent assignee type relates to mentorship within teams using only patent-level observations with teams of more than one inventor. The sample period is from 1845 to 1950. Inventor age gap refers to the age difference between the youngest inventor and the oldest inventor on the team. Company assignee is an indicator that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only. County refers to the inventor's location at the time of patent grant and is based on the first inventor's location for patents with multiple inventors. Tech class refers to the patent's CPC technology class. *t*-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the county level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Dependent variable	Inventor age gap ≥ 5 years	Inventor age gap ≥ 10 years
	(1)	(2)
Company assignee	0.030***	0.020***
	(5.14)	(3.68)
County × Time FE	Yes	Yes
Tech class × Time FE	Yes	Yes
Number of Obs.	182,018	182,018
Adjusted R^2	0.033	0.028
Sample	Census Matched Sample (1845-1950)	

Table 7: Are firms more likely to produce important patents?

This table reports the regression results on the relation between company assignee and patent importance using patent-level observations. The sample period is either the full sample from 1836 to 2018 or the census matched sample from 1845 to 1950. Panel A presents results based on a measure of patent quality and Panel B presents results based on a measure of radical innovation. Patent importance comes from Kelly, Papanikolaou, Seru, and Taddy (2023), which is defined as the ratio of 10-year forward patent text similarity (i.e., the aggregate patent text similarity between the patent and all patents granted in the next ten years) to the 5-year backward patent text similarity (i.e., the aggregate patent text similarity between the patent and all patents granted in the prior five years). Breakthrough patent is an indicator that equals one (zero otherwise) if the patent falls in the top 10% of the unconditional distribution of the patent importance measure. Company assignee is an indicator that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only. Team patent is an indicator that equals one if the patent has more than one inventor and zero if the patent has only a single inventor. County refers to the inventor's location at the time of patent grant and is based on the first inventor's location for patents with multiple inventors. Tech class refers to the patent's CPC technology class. t-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the county level. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Panel A: Measure of Patent Quality

Dependent variable	Patent importance	Patent importance	Patent importance	Patent importance
	(1)	(2)	(3)	(4)
Company assignee	0.028***	0.024***	0.016***	0.014***
	(18.60)	(16.24)	(8.30)	(7.09)
Team patent		0.008***		0.001
		(11.64)		(1.35)
Company assignee × Team patent		0.007***		0.012***
		(7.10)		(5.87)
County × Time FE	Yes	Yes	Yes	Yes
Tech class × Time FE	Yes	Yes	Yes	Yes
Number of Obs.	5,267,855	5,267,855	1,284,333	1,284,333
Adjusted R^2	0.504	0.504	0.501	0.501
Sample	Full Sample (1836-2018)		Census Matched Sample (1845-1950)	

Panel B: Measure of Radical Innovation

Dependent variable	Breakthrough patent	Breakthrough patent	Breakthrough patent	Breakthrough patent
	(1)	(2)	(3)	(4)
Company assignee	0.032***	0.027***	0.023***	0.021***
	(18.51)	(16.20)	(8.66)	(7.69)
Team patent		0.010***		0.001
		(11.84)		(1.06)
Company assignee × Team patent		0.009***		0.011***
		(5.77)		(4.53)
County × Time FE	Yes	Yes	Yes	Yes
Tech class × Time FE	Yes	Yes	Yes	Yes
Number of Obs.	5,267,855	5,267,855	1,284,333	1,284,333
Adjusted R^2	0.129	0.501	0.129	0.129
Sample	Full Sample (1836-2018)		Census Matched Sample (1845-1950)	

Appendix A: Variable definitions

Variable	Description
Panel A: Inventor-level variables	
Male	An indicator variable that is equal to one if the inventor is male, and zero otherwise.
Age at time of patent	The age (in years) of an inventor at the time of patent grant.
Immigrant inventor	An indicator variable that is equal to one if the inventor is a first-generation immigrant, and zero otherwise.
Minority inventor	An indicator variable that is equal to one if an inventor belongs to one of the ethnic minority groups (i.e., Hispanic, Black, Asian or Pacific Islander, and American Indian), and zero otherwise.
% of immigrants from the same origin country as the inventor in the county	The percentage of individuals from the inventor's county that came from the same original native country as the immigrant inventor.
% of minorities from the same minority group as the inventor in the county	The percentage of individuals from the inventor's county belonging to the same ethnic minority group as the inventor.
Panel B: Patent-level characteristics	
Company assignee	An indicator variable that equals one if a patent's assignee is a company and zero if the patent's assignee is composed of individuals only.
Independent assignee	An indicator variable equal to one if the patent's assignee/s are only composed of individuals, and zero otherwise.
Team patent	An indicator variable that is equal to one if the patent is developed by more than one inventor, and zero if the patent only has a single inventor.
Number of inventors	A count of the number of inventors listed in the patent document.
Patent importance	Following Kelly et al. (2021), this is defined as the ratio of 10-year forward patent text similarity (i.e., the aggregate patent text similarity between the patent and all patents granted in the next ten years) to the 5-year backward patent text similarity (i.e., the aggregate patent text similarity between the patent and all patents granted in the prior five years).
Breakthrough patent	An indicator variable that is equal to one if the patent falls in the top 10% of the unconditional distribution of the patent importance measure, and zero otherwise.
Panel C: Team-level characteristics	
Mixed immigrant team	An indicator variable that is equal to one if the team has at least one immigrant inventor and at least one non-immigrant inventor, and zero otherwise.
Mixed minority team	An indicator variable that is equal to one if the team has at least one minority inventor and at least one non-minority inventor, and zero otherwise.
Mixed immigrant/minority team	An indicator variable that is equal to one if the team is classified as a mixed immigrant team or a mixed minority team, and zero otherwise.
Number of different states in a team	The count of distinct U.S. states where the inventors on a team are located.
Number of different counties in a team	The count of distinct U.S. counties where the inventors on a team are located.
All immigrant team	An indicator variable that is equal to one if the inventor team is comprised only of immigrant inventors, and zero otherwise.
All minority team	An indicator variable that is equal to one if the inventor team is comprised only of minority inventors, and zero otherwise.

Variable	Description
Panel C: Team-level characteristics (cont.)	
All immigrant/minority team	An indicator variable that is equal to one if the inventor team is classified as an all immigrant team or an all minority team, and zero otherwise.
Same state team	An indicator variable that is equal to one if all the patent's inventors come from the same state, and zero otherwise.
Same county team	An indicator variable that is equal to one if all the patent's inventors come from the same county, and zero otherwise.
Age gap	The age difference between the youngest inventor on the team and the oldest inventor on the team.