Note /Caution to Reader: Thank you for your interest in this piece, which is very much a work-in-progress. Those uninterested in methods can skip Part II and skim Part III. The first case study in Part IV is also under construction. If any, the novel parts of the paper can be found in the Introduction, Part I, and first case study in Part IV. Please do not cite quote or cite the data in Part IV as its still being refined.

INNOVATORS

By Colleen Chien[1]

ABSTRACT

Patents are increasingly being used by economists and social scientists to track innovators that patent, not just patented innovation. These uses are being driven by the relevance of who is innovating and where to a myriad of current debates, as well as the enhanced availability of large patent datasets. To contextualize this work, this article provides a history of patent law and patenting from an innovator- rather than innovation-centric perspective, and describes some of the cautions that must be taken when using patent data to increase the study innovators that patent, not just patented innovation. To extend it, this article describes and discloses several sources and, in an accompanying appendix, tools for tracking the demographic and economic characteristics of innovators. In a first case study, the application of these tools reveals a thus far surprising and overlooked source of Asian (Chinese and Korean) advantage in innovation including in artificial intelligence - its women. A second case study reveals one reason why, despite seemingly dramatic changes in the patentability of diagnostic innovation, innovation remains surprisingly robust - the large role of nonprofit relative to for-profit entities in advancing diagnostic innovation. A third mini-case study evaluates the effect of the introduction of a new set of policies introduced by the America Invents Act of 2012 intended to increase participation of the smallest innovators in the patent system. Collectively, they illustrate the various ways in which research focused on *innovators* can enrich and improve policymaking and decision-making about *innovation.*

---

[1] Professor, Santa Clara Law School, 2013-2015 Senior White House Advisor, Innovation and Intellectual Property. I thank Rochelle Dreyfuss, Jeanne Fromer, Eric Goldman, Pratheepan Gulasekaram, David Schwartz, Tom Cotter and Michael Risch and audiences at Cardozo Law School, NYU Law School, San Diego Law School, Case Western Law School, Texas Law School, Colorado Law School, and Santa Clara Law for their comments on earlier drafts. I thank Jiun Ying Wu, Jerome Ma, Marinna Radloff, and Maria Sokova for excellent research assistance and Kristen Forrester for administrative help. I thank NamePrism, Bill Kerr, Lisa Cook, and the USPTO Chief Economist Office for help with data. Thanks always to Dirk Calcoen. A codebase will be made available with the tools used to generate the analysis.

INTRODUCTION

The patent system exists to promote innovation. By offering limited rights to exclude others from practice of the invention, the patent system "add[s][] fuel to the fire of genius."[2] Consistent with the primacy of this utilitarian purpose,[3] scholars of the patent system have generally prioritized the advancement of patented innovation over the advancement of innovators that patent. In this article I use the term "innovators" to refer to the narrow group of innovators that creates or own patents.[4] Doctrinal scholarship, for example, is largely concerned with how to strike the right balance between innovation and competition, not to enhance the welfare of inventors or consumers as such.[5] Empirical studies aren't much better— subject to some notable exceptions,[6] most quantitative studies are based on raw patent counts, citations, and metrics[7] that differentiate, if at all, based on industry[8] or technology,[9] rather than innovator type.

The focus on innovation, not innovators, is understandable. Patents are available for inventions that are novel, nonobvious, and adequately described—all qualities of the underlying invention, not the inventor. Unlike other American institutions like voting, patent eligibility has never explicitly depended, for example, on an inventor's gender.[10] Patent counts are routinely used and relied upon by governments and scholars as measures of innovation that, while imperfect, are easier to come by than data about inventors.[11] While copyrighted works such as

---

[2] Abraham Lincoln, Lecture on Discoveries and Inventions (Feb. 11, 1859), *in* THE POLITICAL THOUGHT OF ABRAHAM LINCOLN 112, 121 (Richard N. Current ed., 1967).

[3] *See, e.g.*, David S. Olson, *Taking the Utilitarian Basis for Patent Law Seriously: The Case for Restricting Patentable Subject Matter,* 82 TEMP. L. REV. 181, 182 (2009) ("There is widespread agreement that the reason we have a patent system is utilitarian."); Jeanne C. Fromer, *Expressive Incentives in Patent Law,* 98 VA. L. REV. 1746, 1750-51 ("The Supreme Court, Congress, and many legal scholars consider utilitarianism the dominant purpose of American . . . patent law.").

[4] William Fisher, Theories of Intellectual Property 2, 8 (2001) (unpublished manuscript), *available at* https://cyber.harvard.edu/people/tfisher/iptheory.pdf (arguing that the dominant, incentive-based model of intellectual property discounts the importance of social and communal goals that advance the interests of users and creators, not just owners of intellectual property); *see also* Betsy Rosenblatt, *Intellectual Property's Negative Space: Beyond the Utilitarian* 40 FLA. ST. U. 411 (2011) (describing the underappreciated significance of non-utilitarian theories including labor-desert, personality, and distributive justice theories to creation and innovation).

[5] With some exceptions, as discussed in Part I.B.

[6] Detailed in Part I, *infra*.

[7] Described in Part II, *infra*, e.g., at Table 2A.

[8] See studies cited in Part II (describing efforts to match patents to industries—most notably by the NBER Patent Citations File, which categorized patents into six economically-relevant technology categories and 37 sub-categories).

[9] *See, e.g.*, Alberto Galasso & Mark Schankerman, *Patents and Cumulative Innovation: Causal Evidence from the Courts* 24 (Nat'l Bureau of Econ. Research (NBER), Working Paper No. 20269, 2014) (contrasting cumulative and discrete innovation based on the complexity of end product; in cumulative or complex innovation fields "new products embody numerous patentable elements" and pioneering and incremental innovation); John R. Thomas, *The Question Concerning Patent Law and Pioneer Inventions,* 10 BERKELEY TECH. L.J. 35 (1995) (describing the differential treatment under the law accorded to "revolutionary" pioneer inventions as opposed to more incremental inventions).

[10] Though, as Part I explores, it has, over time, systematically excluded black Americans and non-white foreigners.

[11] Described in Parts I–II.

music, books, and movies are often intimately connected to the artist or author, patented innovations are much less likely to be tied to individual inventors.[12]

This Article supports reconceptualization of the patent system as a way to understand and promote innovators that patent, not just patented innovation.[13] Although the role of innovators is instrumental, rather than personal, innovation only advances at the behest of "authors and inventors," as enshrined in the Constitution. Who is innovating and where innovation occurs are relevant to a myriad of current debates, including the inclusion of women and minorities in innovation, high-skilled immigration, and national competitiveness.[14] As this study demonstrates, computational, artificial-intelligence powered classifiers, when applied to open patent records, can estimate details about innovators that patent such as likely race and gender, fields that may otherwise be encumbered by privacy protections. Indicia of company number of employees and revenue, which are often available only for public companies, are also available in the patent record. Studying innovators is also relevant to patent policy-making, because a distinct strand of patent policy has always been focused on innovators, not only innovations,[15] and greater scholarly engagement could improve their performance. Finally, innovator-based theories have explanatory value, as different types of innovators experience the innovation system differently and may require incentives that are calibrated to the needs of, for example, startups or regionally-clustered innovators. As such, there are utilitarian, policy, and theoretical reasons to view the patent system as a way to understand *innovators*, not just *innovation*. These reasons are best understood within the literatures and debates upon which this work builds upon and to which it hopes to contribute.

The first literature concerns the patent system's performance as an inclusive public institution. Influential scholars have held up the early patent system as an example of the type of democratic institution responsible for American prosperity. As Acemoglu and Robinson wrote in their landmark work, *Why Nations Fail*: "[j]ust as the United States in the 19th Century was more democratic politically than most any other nations in the world at the time, it was also more democratic than others when it came to innovation. This was critical to its path to becoming the most economically innovative nation in the world."[16] Recently, however, the patent system has come under fire by prominent commentators as having the opposite effect of widening inequality, and enriching the intellectual property haves at the expense of consumers and intellectual property have-nots.[17] Though economic historians have done important work on the

---

[12] Personhood interests are also less developed in patents than in copyright, Fromer, *supra* note ___, at 1754 (acknowledging that "personhood theory is less frequently invoked as an explanation for patent law," than copyright law, but also arguing that inventors have overlooked personhood interests in their inventions).

[13] As Part II describes, patented innovation is only a small subset of overall innovation, but a relatively larger share of commercially significant innovation.

[14] Described in Part I.

[15] *Id.*

[16] DARON ACEMOGLU & JAMES A. ROBINSON, WHY NATIONS FAIL: THE ORIGINS OF POWER, PROSPERITY, POVERTY 333 (2013).

[17] *See, e.g.*, JOSEPH STIGLITZ, REWRITING THE RULES OF THE AMERICAN ECONOMY: AN AGENDA FOR GROWTH AND SHARED PROSPERITY __ (2016) (arguing that patent-based rent-seeking is responsible for widening inequality); Dean Baker, *The Upward Redistribution of Income: Are Rents the Story?*, 48 REV. RADICAL POL. ECON. 529 (2016) (arguing that increasing rents in four areas, including patents and copyrights, are responsible for the bulk of the upward redistribution of income in recent decades). This isn't necessarily a new criticism. *See, e.g.*, Shiyuan Pan, Heng-fu Zou, & Tailong Li, *Patent Protection,*

democratizing impact of the early patent system,[18] this work is in need of updating as it precedes the rise and current dominance of industrial and corporate inventing.[19] By describing, then applying novel methods and approaches to patent data from the last four decades, this work addresses how well the patent system continues to live up to democratic, pluralistic ideals—a question with growing relevance as U.S. companies increasingly capture value based on intangible, rather than tangible, assets.

This work also speaks to open questions about the design of patent law and policy. In theory, the patent system is unitary and one size fits all. Discrimination based on "the field of technology"[20] is prohibited under international law,[21]limiting the ability of legislators to design provisions to meet the disparate needs of different industries.[22] Policymakers have much greater latitude to tailor patent law by innovator type than by industry, however.[23] But to do so, they must be able to identify innovator types and discern their prevalence and preferences. Industries that benefit from intellectual property are among those that spend the most on lobbyists;[24] certain innovator groups are also well represented, for example, through the

---

Technological Change and Wage Inequality (China Econ. & Mgmt. Acad., Working Paper No. 437, 2010), http://down.aefweb.net/WorkingPapers/w437.pdf (reviewing the economics literature that explores the contribution of strengthened intellectual property to wage inequality).

[18] Reviewed in Parts II and III (see descriptions of work by Khan, Sokoloff, and Lamoreaux).

[19] Described, e.g., in Lemelson-MIT Program Proceedings, *Historical Perspectives on Innovation and Creativity* (2003), at 19-25 (describing the dominance of corporate, not Edisonian, labs in U.S. innovation starting in the 1920s). *See also* Michael Risch, *Licensing Acquired Patents*, 21 GEO. MASON L. REV. 979, 999 (2014) (arguing that various development have led to a recent growth in the importance and relevance of independent inventors).

[20] Agreement on Trade-Related Aspects of Intellectual Property Rights, art. 27(1), Apr. 15, 1994, 1869 U.N.T.S. 299 (establishing that World Trade Organization ("WTO") countries must offer protection to any invention, whether product or process, without discrimination based on the field of technology) [hereinafter TRIPS]. *But see id.* art. 27(2)-(3) (sanctioning exceptions to patentable subject matter on a variety of grounds including for public order or health, the environment, as well as supporting the exclusion of diagnostic, therapeutic, and surgical methods for the treatment of humans or animals from patentable subject matter).

[21] Despite this, U.S. and European patent law have numerous technology-specific provisions, as described in Colleen V. Chien, *Tailoring the Patent System to Work for Software and Technology Patents* (Nov. 15, 2012), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2176520.

[22] Described, e.g., in Dan L. Burk & Mark A. Lemley, *Patent Crisis and How the Courts Can Solve It*, _____ (arguing that courts are in a better position to adopt the law to the disparate); *see also* FED. TRADE COMM'N, THE EVOLVING IP MARKETPLACE: ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION 10 (2011) (noting the "divergence in the extent and nature of notice problems among industries").

[23] This power to tailor patent law to the needs of specific innovators is not unlimited, however; the national treatment principle, for example, requires that the law treats foreigners as well as it treats domestics. *See* TRIPS art. 3 ("[e]ach [WTO] Member shall accord to the nationals of other Members treatment no less favourable than it accords to its own nationals with regard to the protection of intellectual property").

[24] For example, from 1998 to 2016, "Big Pharma" spent close to $3.5B—more than any other industry—on lobbying expenses, though the specific share of lobbying dollars allocated to patent policy is unknown. Michelle Llamas, *Big Pharma and Medical Device Manufacturers,* DRUGWATCH, https://www.drugwatch.com/manufacturer/ (last visited Oct. 30, 2017). In 2015 and 2016, Pharmaceuticals/Health Products and Electronics Manufacturing & Equipment industries were both among the top five spenders on lobbying. OPENSECRETS.ORG, LOBBYING SPENDING DATA, YEARS 2015 &

university,[25] tech,[26] and independent inventor[27] lobbies. But consistent with public choice theory, special interest lobbies are not necessarily representative of all affected parties. Using the tools described in Part II, the economic significance and size of innovator groups and how various policy interventions would impact disparate stakeholders can more readily be discerned.

This work is also relevant to broader social and academic conversations about inclusion and diversity in innovation. It has long been assumed that technological innovation is driven by value-neutral, measurable outcomes related to technical performance. But recent stories of the pervasive hostility of the tech sector to minorities and women,[28] not unlike the revelations that rocked the financial industry in the 1990s,[29] have shaken to the core the belief that the technology industry functions as a meritocracy.[30] The underrepresentation of women[31] and certain minorities[32] in innovation is well-documented. This work borrows from and extends this literature to facilitate the identification of successes and failures, and develop techniques that can be used to inform and evaluate tech inclusion initiatives.

2016, https://www.opensecrets.org/lobby/top.php?showYear=2015&indexType=i (last visited Oct. 30, 2017).

[25] For a description of the university lobby, its activities, and its influence, see Peter Lee, *Patents and the University*, 63 DUKE L.J. 1, 5 (2013).

[26] Described, e.g., in Issie Lapowskey, *What Tech Giants Are Spending Millions Lobbying For*, WIRED (July 23, 2015, 7:00AM), https://www.wired.com/2015/07/google-facebook-amazon-lobbying/ (noting substantial lobbying attention from Google, Amazon, Facebook, and Apple for patent reform).

[27] Described by, e.g., John R. Allison et al., *Valuable Patents,* 92 GEO. L.J. 435, 468 (2004) (discussing the influence of the small inventor lobby); Christopher A. Cotropia, *The Individual Inventor Motif in the Age of the Patent Troll,* 12 YALE J.L. & TECH. 52 (2009) (describing the activities of the individual inventors' lobby).

[28] *See, e.g.*, ALLISON SCOTT ET AL.,TECH LEAVERS STUDY, KAPOR CENTER FOR SOCIAL IMPACT (Apr. 17, 2017), http://www.kaporcenter.org/wp-content/uploads/2017/04/KAPOR_Tech-Leavers-Final.pdf.

[29] For an overview, see Louise Roth, *Women on Wall Street: Despite Diversity Measures, Wall Street Remains Vulnerable to Sex Discrimination Charges,* ACAD. MGMT. EXECUTIVE, (Feb. 2007) (describing how high-profile charges of sexual harassment and costly lawsuits in the finance industry in the 1990s catalyzed the implementation of industry-wide policies, though of questionable effectiveness).

[30] Joan C. Williams, *Hacking Tech's Diversity Problem,* HARV. BUS. REV., (Oct. 2014), https://hbr.org/2014/10/hacking-techs-diversity-problem (describing as "[a] key feature of the tech culture—the shared belief that it's a meritocracy."); Henry M. Chesbrough & Melissa M. Appleyard, *Open Innovation and Strategy,* 50 CAL. MGMT. REV.57, 69 (2007) (describing the open innovation community as conceiving of itself "operating as a meritocracy.").

[31] For an overview of this literature, see Adams Nager et al., *The Demographics of Innovation,* INFORMATION TECHNOLOGY & INNOVATION FOUNDATION 8-9 (Feb. 2016), *available at* http://www2.itif.org/2016-demographics-of-innovation.pdf (describing the underrepresentation of women in the STEM workforce, high-growth entrepreneurship and a variety of innovation competitions) [hereinafter ITIF]. Some work has also been done to consider whether or not patent law contains implicit gender biases, see, e.g., Dan L. Burk, *Diversity Levers*, 23 DUKE J. GENDER L. & POL'Y 26 (2015) (arguing that the nonobviousness standard in patent law, in practice, is not gender neutral and contributes to implicit social biases that raise significant barriers for women to successful creativity and innovation), or reflects masculine theories of capitalism, see, e.g., Laura A. Foster, *Patents, Biopolitics, and Feminism: Locating Patent Law Struggles Over Breast Cancer Genes and the Hoodia Plant,* 19 INT'L J. CULTURAL PROP. 371 (2012) (suggesting critical inquiries for the formulation of a feminist analysis of patent law).

[32] *Id.*, ITIF, *supra* note ___, at 10; *see also* descriptions in Parts I.B. and IV of work by Bell et al., Cook & Kongcharoen, and others.

This Article proceeds as follows: Part I makes the case for reconceptualizing the patent system as a set of policies and information about *innovators*, not just *innovation*. Studying and promoting innovators who patent, rather than just patented innovations, can strengthen and improve innovation outcomes on a number of dimensions.[33] By explicitly acknowledging the largely neglected portions of patent law that are innovator-, not just innovation-driven, from the early exclusion of foreigners and non-whites from patenting to the contemporary promotion of the smallest inventors, this Part exposes the need for greater academic attention to the impacts and efficacy of this strand of the law.

Part II addresses *how* to use the patent record to study innovators. Techniques for studying innovation through the patent record are well-developed, and hand-coded methods have been applied, for example, to the identification of different types of patent litigants and contests.[34] However, the time is ripe for the broader consideration of ways to access *innovator* information at scale, with the recent release by the USPTO of high-quality data and development of new computational tools. Understanding how the newly available data are constructed, what they do and don't measure, and their strengths and weaknesses are essential to their use.

By accessing the sources and applying the tools described in Part II, Part III provides a descriptive context for research on innovators that addresses who invents, who owns inventions, and where is innovation happening? Part IV explores several directions in innovator research based on applying the approaches described in Parts I and II to specific research and policy issues. It reports on three mini-case studies that show how focusing on innovators can improve innovation. The first case study evaluates a recent patent policy change intended to increase participation of the smallest innovators in the patent system. The second case study considers the disparate trajectories of diagnostics innovators of different sizes following changes in the law of patentable subject matter.

PART I: WHY INNOVATORS

If the purpose of patent law is to promote innovation, why should it matter who is innovating? This Part explores several reasons. First, though the patent system is typically thought of as source of information about patented innovation, it is also a rich source of information about inventors and owners of patents, or "innovators" that patent as demonstrated by disparate strands of historical, economic, and social science literature. Second, though largely overlooked, American patent law has long been the law of innovators, not just innovation, and this strand of policymaking deserves greater academic scrutiny and evaluation. Finally, who is innovating matters because different innovators use and experience of the patent system differently, whether before the Patent Office, in court, or otherwise. Using the patent record to access the experiences of disparate groups of innovators can improve public and corporate innovation policymaking.

---

[33] *See* Part I, *infra*.

[34] John R. Allison and his colleagues, for example, created a taxonomy of 12 types of patent asserters in *Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents*, 158 U. Pa. L. Rev. 1, 10-11 (2009) which served as the basis for the Stanford NPE litigation dataset created by Shawn Miller et al, which aggregates numerous hand-coded datasets, each generally detailing hundreds or thousands of parties. Shawn P. Miller et al., Introduction to the Stanford NPE Litigation Dataset (Oct. 23, 2017) (unpublished manuscript), *available at* https://law.stanford.edu/wp-content/uploads/2017/10/Introduction-to-the-Stanford-NPE-Litigation-Dataset-10.23.2017.pdf.

A. An Improved Understanding of Innovators Through the Patent Record Can Lead to The Improved Cultivation and Promotion of Innovation

Behind every innovation is an innovator or group of innovators. Highly-talented individuals have an outsized impact on innovation and the trajectory of history,[35] and as such, the risks associated with misallocating resources to talent, as explored through a burgeoning economics literature,[36] are great. As such, it is not surprising that promoting innovators is a cornerstone of many non-patent innovation policies.

For example, what if, as Facebook's Mark Zuckerberg recently asked publicly, he hadn't been born into a financially stable, middle-class family that afforded him with the "freedom to fail"?[37] Albert Einstein's family owned a manufacturing company and provided with him a high-quality education in Europe;[38] but what if instead, they needed young Albert to work to support the family? Celik has estimated that doing a better job of capturing the talent of "lost Einsteins," could grow the economy by a rate of 10%,[39] an impossibly high number but one that illustrates the stakes. If women, minorities, and children from low-income families were to invent at the same rate as white men from high-income families, there would be four times as many inventors in America as there are today, Chetty and his colleagues have found.[40] Another set of policy questions lies at the intersection of innovation and human resources. Broad-based access to high-quality science, technology, engineering, and mathematics (STEM) education is generally viewed as a building block of a robust innovation ecosystem.[41] So, to many, STEM is clearing a path for immigrants, particularly those with specialized technical skills and entrepreneurs, to "help propel the innovation economy."[42] These and related policy areas convey the importance of cultivating innovators to cultivating innovation.

But while patents have long been regarded as important, albeit highly imperfect,[43] sources of information about innovation,[44] less attention has been paid to patents as sources of

---

[35] Described, e.g. in Tim Sullivan, *Entrepreneurs, Economic Growth, and the Enlightenment,* HARV. BUS. REV. (August 10, 2015) (reviewing the literature on the outsized impact of star scientists and entrepreneurs on innovation, and the importance of innovation to economic growth).
[36] *See, e.g.*, Murat Alp Celik, *Does the Cream Always Rise to the Top* 1-2 (2015) (unpublished Ph.D dissertation, University of Pennsylvania) *available at* https://www.tse-fr.eu/sites/default/files/TSE/documents/sem2016/jobmarket/jmp_celik.pdf.
[37] Catherine Clifford, *Mark Zuckerberg: Success comes from 'the Freedom to Fail*, CNBC May 25, 2017, https://www.cnbc.com/2017/05/25/mark-zuckerberg-on-success-billionaires-should-pay-you-fail.html
[38] Celik, *supra* note __ at 1.
[39] *Id.*
[40] Alex Bell et al., *The Who Becomes an Inventor in America? The Importance of Exposure to Innovation,* 43 (December 2017) available at http://www.equality-of-opportunity.org/assets/documents/inventors_paper.pdf.
[41] As described in NAT'L ECON. COUNCIL & OFF. OF SCI. & TECH. POL'Y, EXEC. OFFICE OF THE PRESIDENT, A STRATEGY FOR AMERICAN INNOVATION (Oct. 2015), *available at* https://obamawhitehouse.archives.gov/sites/default/files/strategy_for_american_innovation_october_2015.pdf.
[42] *Id.* at 3, 29 (describing immigration—in particular, of high-skilled workers and entrepreneurs—as a building block of the American innovation ecosystem).
[43] Part II addresses the weaknesses of patent data for tracking innovation, starting with the fact that many innovations are not patented. For an overview of the distinction between invention and innovation, see

data about innovators. This is in part because, while it has long been true that patents contain the names, locations, and other details about the subset of innovators that patent, accessing this information, and translating it into meaningful insights has historically been difficult and time-intensive, requiring extensive manual collection, cleaning, and processing.

But in recent years, many of the barriers to the use of patent records for data on innovators have disappeared. Extensive efforts have gone into the normalization of innovator names,[45] making it easier to attribute patenting behavior to particular inventors and assignees. The U.S. and other jurisdictions have, in the last decade, made patent data available in bulk format pertaining not only to the patent itself but transactions of the patent.

Patent records have numerous advantages over alternate sources of innovator data. Unlike surveys, patent records cover all, not just a sample, of those with the particular trait of having filed for a patent, and include penalties for misrepresenting legal facts.[46] Because patent records are part of the public record, few privacy or proprietary barriers stand in the way. Revenue and employee number information, for example, is not generally available for private companies. But when companies below a certain size apply for patents, they must declare their size below a threshold to receive a discount. This makes it possible to identify small firms with some degree of confidence through the patent record.[47] From inventor names, which are part of the public record, recently-developed name-based classifiers can be used to infer with a reasonably high level of confidence the ethnicity and gender of inventors.[48] From the city and state listed on a patent, one can tell whether the invention was developed in a foreign or domestic, or in a rural or urban setting. The patent record also reveals numerous details about

---

Jan Fagerberg, *Innovation: A Guide to the Literature* 3-5, (Ctr. for Tech., Innovation & Culture, Univ. of Oslo, Oct. 12, 2003) *available at* https://smartech.gatech.edu/bitstream/handle/1853/43180/JanFagerberg_1.pdf?sequence=1&isAllowed=y. For all the reasons described *infra*, patents are an imperfect measure of even new ideas.

[44] Since Jacob Schmookler's pioneering work assigning patents to industries a half century ago, many have sought to glean economic and technical insights from patents. Jacob Schmookler, *Economic Sources of Inventive Activity*, 22 J. ECON. HIST. 1 (1962). For an overview, see Bronwyn Hall et al., *NBER Patent Citations Data File: Lessons, Insights and Methodological Tools*, (NBER, Working Paper No. 8498, 2001) (describing efforts by Scherer and Griliches to connect patents to industries and firms as well the NBER patent citations file which the authors produced in order to make available en masse, and at firm- and industry-level patent citations) [hereinafter NBER Patent Citations] and Alan C. Marco et al., *The USPTO Historical Patent Data Files* (USPTO, Working Paper No. 1, 2015) (describing extensions to the NBER citations data file completed by the authors).

[45] Historically, the lack of strict naming protocols for inventor and owner names has made it difficult to trace the patents associated with any single innovator. For example, the corporation IBM has been presented on the face of various patents as International Business Machines, I.B.M., IBM Corp. Int'l Business Machines, and so on, with one data provider estimating that there are over 2,000 names for IBM alone. However, the USPTO's considerable efforts to normalize names has resulted in the production of clean datasets for analyzing inventors. For an overview of efforts to normalize data and why they are important, see Manuel Trajtenberg et al., *The "Names Game": Harnessing Inventors, Patent Data for Economic Research*, 93/94 ANNALS ECON. & STAT. 79 (2009).

[46] Such as company size, see discussion *infra* in Part II.

[47] See description of small entity and micro-entity status *infra* and in Part II.

[48] As illustrated in Part II. *See also* C. Fritz Foley & William R. Kerr, *Ethnic Innovation and U.S. Multinational Firm Activity*, 59 MGMT. SCI. 1529 (2013) (determining the ethnicity of inventors based on their last names, using the techniques described *infra*); Lisa D. Cook & Chaleampong Kongcharoen, *The Idea Gap in Pink and Black* (NBER, Working Paper No. 16331, 2010) (identifying African-American and women inventors using various methods).

the inventive process, including the number of inventors that contributed to an invention[49] and whether the invention was developed within or outside of a corporate setting based on who owned it at the time of patenting. Studying public records can also convey whether or not the invention was licensed, litigation, or re-assigned, which are indicia of commercialization,[50] and how such characteristics vary by the inventor's geography.[51] Such innovator traits, about the owner as well as the inventor of the patent, are relevant to policies described above.

Another advantage of patent records for innovators, not only innovation, is that they capture the behavior of all who patent— small and large, non-profit and for-profit, and foreign and domestic—and therefore support the testing of findings about innovators across large populations and time. For example, a 2011 study that considered research and development teams in Spain found that companies with greater gender diversity were more likely to introduce "radical new innovations during the studied period."[52] A 2013 study of London firms similarly found that companies with diverse management were more likely to introduce new product innovations than were those with homogenous leadership.[53] However, by focusing on particular swaths of firms, these studies address whether or not their findings were generalizable beyond Spanish and British contexts. Patent data can address these deficiencies by enabling observed trends to be tested across the different settings of patenting, as well as to support adjustments based on quality, according to metrics like a patent's subsequent citation (also known as a patent's "forward citation").[54] Independent studies have found, for example, that patents by mixed-sex teams are more valuable across countries, according to this metric, than male-only patent teams.[55]

Yet another strength of patent records is their granularity, as each patented invention is associated with one or more technology classes, inventors, owners, locations, and dates. These details allow patent data to be further subdivided into various groupings based on, for example, technology, geography, gender, ethnicity, or time. They also support combining patent data with other sources of data.[56] Several economists have developed novel approaches to glean insights

---

[49] *See, e.g.*, Dennis Crouch, *The Changing Nature of Inventing: Collaborative Inventing*, PATENTLY-O (July 9, 2009), http://patentlyo.com/patent/2009/07/the-changing-nature-inventing-collaborative-inventing.html (noting the rising number of patents that name multiple inventors).

[50] For an overview of the commercial significance of these events, see Colleen V. Chien, *Predicting Patent Litigation,* 90 TEX. L. REV. 283 (2011).

[51] For similar approaches, see, e.g., Naomi R. Lamoreaux & Kenneth L. Sokoloff, *Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries* __ (Nat'l Bureau of Econ. Research, Working Paper No. 98, 1997) (comparing patenting and patent trading rates by regions of the United States, as well as rates of independent invention and corporate assignment).

[52] Cristina Diaz-Garcia, Angela Gonzalez-Moreno & Francisco Jose Saez-Martinez, *Gender Diversity within R&D Teams: Its Impact on Radicalness of Innovation*, 15 J. INNOVATION: ORG. & MGMT. 149 (2011).

[53] Max Nathan & Neil Lee, *Cultural Diversity, Innovation, and Entrepreneurship: Firm-level Evidence from London*, 89 J. ECON. GEOGRAPHY 367 (2013).

[54] See Hall and Harhoff, *supra* note __ at 20 for a review of the literature on the relationship between a patent's forward citations and economic value.

[55] Cook & Kongcharoen, *supra* note __, at 3; Catherine Ashcraft & Anthony Breitzman, *Who Invents IT? Women's Participation in Information Technology Patenting, 2012 Update*, NAT'L CTR. FOR WOMEN & INFO. TECH. 1, 4 (last visited Aug. 7, 2017).

[56] For example, academic researchers with sufficient permissions can access detailed demographic and income data for inventors. Public LinkedIn profiles can provide employment and interim location data for individual inventors.

about innovators using patent data, as described in Part II, but much work remains to be done, both to explore the relationship between innovators—and to support them through private and public interventions—and innovation, and to translate descriptive insights, including the results of experimentation, into policy suggestions. The sourcing of innovator metrics from patent data and federation of this data with other sources must be done carefully and with full awareness of their strengths and weaknesses. Fortunately, many of the commonly understood cautions that apply to the use of patents as useful though imperfect innovation metrics also extend to patents as innovator metrics. Both are discussed further in Part III.

B. Patent Law Is the Law of Innovators as Well as the Law of Innovation

Another reason to care about who is innovating is because policymakers have always cared. From the origins of the U.S. patent system to the present, the law has both encouraged and discouraged the participation of certain classes of innovators. At different times, the law has also singled out particular types of inventors, including individuals, for special treatment. Taken together, these laws and policies reinforce the degree to which patent law has and continues to be the law of innovators, not just innovation.

The Patent Act of 1790 authorized anyone who invented or discovered "any useful art, manufacture, engine, machine, or device, or any improvement therein" to apply for a patent.[57] The first Act was remarkably inclusive for its time — in contrast to naturalization, which was reserved for "free White Persons"[58]—"any person or persons" could apply for a patent.[59] All who did received the same rights, unlike the discounting of slaves to "three-fifths of ...Persons" for purposes of taxation and representation enshrined in the U.S. Constitution.[60] In contrast to suffrage, which was not guaranteed for women until 1920,[61] "he, she, or they" could apply for a patent.[62] But the exceptional openness of the earliest patent law was short lived. The Patent Act of 1793 restricted eligibility to patent to U.S. citizens.[63] This meant that foreigners, slaves, and non-white immigrants, that is, those who were not "free White persons" eligible under the 1790 Immigration and Naturalization Act for citizenship, could not apply for or get patents, though women still could.

---

[57] Patent Act of 1790, ch. 7, § 1, 1 Stat. 109 [hereinafter 1790 Act].

[58] Naturalization Act of 1790, ch. 3, § 1, 1 Stat. 103. This excluded naturalization of Asians, American indians, and free black immigrants. According to Haney-Lopez, this racial prerequisite to citizenship remained in force until 1952. IAN HANEY-LOPÉZ, WHITE BY LAW: THE LEGAL CONSTRUCTION OF RACE 1 (New York Univ. Press rev. ed. 2006).

[59] 1790 Act § 1.

[60] U.S. CONST. art. II, § 2.

[61] Through the ratification of the 19th Amendment to the Constitution, which states that "[t]he right of citizens of the United States to vote shall not be denied or abridged by the United States or by any State on account of sex."

[62] 1790 Act § 1. Mary Kies is believed to be the first woman to be granted a U.S. patent. Erin Blakemore, *Meet Mary Kies, America's First Woman to Become a Patent Holder*, SMITHSONIAN: SMARTNEWS, (May 5, 2016). http://www.smithsonianmag.com/smart-news/meetmary-kies-americas-first-woman-become-patent-holder-180959008/.

[63] Patent Act of 1793 Act, ch.11, § 1, 1 Stat. 318-23.

The preferences for free, white, male, domestic U.S. innovators persisted until about 1860.[64] Slave owners exploited the law to their advantage as Eli Whitney became famous based on a cotton gin now attributed to a slave named Sam, and, according to accounts, the "McCormack" reaper actually benefited greatly from the contributions of a slave named Jo Anderson.[65] The rights of married women[66] and blacks[67] to obtain patents were on uncertain footing for most of the first century of the patent system.[68] Until as recently as 2011, foreign innovators have held fewer rights—whether with respect to the ability to patent, [69] the ability to pursue certain remedies without a U.S. industry,[70] or the consideration of their activities abroad[71]—than their domestic counterparts.

In addition to policies of *exclusion*, a distinct line of *inclusive* patent policies further illustrates the degree to which U.S. patent laws have been innovator- not only innovation-driven. For example, to facilitate participation by rural inventors, the early U.S. patent system allowed for patenting by mail.[72] Features like relatively low fees[73] and the award of patents based on merit to the inventors of original ideas rather than based on patronage[74] also contributed to the "democratization" of the U.S. system.[75] This lineage of *inclusive* patent policy continues to this day. In 1982, Congress introduced fee discounts for small, non-profit, and

---

[64] For the progression of the laws, see Appendix, Table A.
[65] *See* Brian L. Frye, *Invention of a Slave,* ___ SYRACUSE L. REV. ____ (forthcoming) (unpublished manuscript *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2918085, at 4, nn.19-20 and accompanying cites).
[66] *See, e.g.*, *Fetter v. Newhall*, 17 F. 841, 843 (C.C.S.D.N.Y. 1883) (confirming that "minors, married women, and others suffering from a legal disability" were eligible to patent).
[67] For citations, see Appendix, Table A.
[68] Though both women and black inventors managed to get patents during this time, see Frye, *supra* note ____ (describing antebellum patenting by black Americans); B. Zorina Khan, *"Not for Ornament": Patenting Activity by Nineteenth-Century Women Inventors*, 31 J. INTERDISC. HIST. 159 (2000) (providing a history of early patenting). Mary Kies is believed to be the first woman to be granted a U.S. patent, on May 15, 1809. Blakemore, *supra* note ___.
[69] Detailed in Appendix, Table A, at description of the Patent Acts of 1793, 1800, 1832, and 1842.
[70] *Id.*, at description of 1930 Tariff Act.
[71] *Id.*, at description of 1952 Patent Act, TRIPS Agreement, and Leahy-Smith America Invents Act.
[72] B. Zorina Khan & Kenneth L. Sokoloff, *Patent Institutions, Industrial Organization and Early Technological Change: Britain and the United States, 1790-1850 in* TECHNOLOGICAL REVOLUTIONS IN EUROPE: HISTORICAL PERSPECTIVES (Maxine Berg & Kristine Bruland eds., Edward Elgar 1998).
[73] PETER DRAHOS, THE GLOBAL GOVERNANCE OF KNOWLEDGE: PATENT OFFICES AND THEIR CLIENTS 99-109 (2010) (describing U.S. fees as set below U.K. fees at the outset, in 1790, and lower than most European countries, for the first half of the 19th century). *Accord* B. ZORINA KHAN, THE DEMOCRATIZATION OF INVENTION: PATENTS AND COPYRIGHTS IN AMERICAN ECONOMIC DEVELOPMENT, 1790-1920 29 (Cambridge Univ. Press 2005).
[74] As was prominent in Britain at the time of the founding of the U.S., as described in KLAUS BOEHM & AUBREY SILBERSTON, 1 THE BRITISH PATENT SYSTEM: ADMINISTRATION 14 (Cambridge Univ. Press 1967) (describing, during the reign of Queen Elizabeth, a tradition of "the Crown embark[ing] upon a policy of granting exclusive privileges...[which] amounted to a comprehensive industrial policy which it has been suggested was intended to produce Crown participation in, and control of, industry").
[75] As described in Khan & Sokoloff, *supra* note ___, at 292-313 .

individual inventors[76] to level the playing field. Since at least 2006, the Patent Office has offered several ways to prioritize patent applications including based on the age or the health of the applicant.[77] In 2011, Congress explicitly empowered the USPTO to "recognize the public interest in continuing to safeguard broad access to the United States patent system,"[78] lowered fees for,[79] and developed other measures to support the smallest inventors.[80] It created regional offices of the USPTO in Detroit, Dallas, Denver, and San Jose to offer services across the country, not just in Alexandria, Virginia.[81]

In addition to these exclusionary and inclusionary policies, patent law and policy has at times singled out certain classes of innovators for special treatment — in recent decades, universities and patent assertion entities (PAEs) in particular. For example, academic research and development (R&D) is funded, incentivized, and structured in ways that differ significantly from corporate R&D. Accordingly, the courts have historically distanced patent law from academic science in the implementation of several patent doctrines.[82] As academic inventing has become more commercial, Peter Lee has argued, courts have increasingly refused to give preferential treatment to universities.[83] Congress, however, has continued to do so. The 2011 America Invents Act includes a "University Exception" that immunizes university patents from defenses to infringement based on "prior user rights," in effect strengthening university patents relative to others.[84] Following a last-minute amendment to the law, the Act also entitles universities to the same deep discounts off its patent fees that the smallest entities enjoy,[85] even

---

[76] Act of Aug. 27, 1982, Pub. L. No. 97-247, § 1, 96 Stat. 317. For additional history about the introduction of reduced fees, see Jeff A. Ronspies, Comment, *Does David Need a New Sling? Small Entities Face a Costly Barrier to Patent Protection*, 4 J. MARSHALL REV. INTELL. PROP. L. 184 (2004).

[77] 37 C.F.R. § 1.102(c)(1) (2016) ("A petition to make an application special may be filed...if the basis for the petition is: (1) The applicant's age or health.").

[78] Leahy-Smith America Invents Act § 28, Pub. L. No. 112-29,125 Stat. 284, 339 (2011) [hereinafter 2011 America Invents Act].

[79] *Id.* §10(b)-(g), 125 Stat. at 316-18 (2011) (codified at 35 U.S.C. § 123) (establishing micro-entity fees)

[80] *Id.* §§ 3(a)(2) (requiring the Small Business Administration to conduct a study of the Act's impact on small businesses, 28 (requiring the establishment of an ombudsman to provide "support and services relating to patent filings to small business concerns and independent inventors"), 31 (requiring the USPTO to conduct a study of how the federal government "can best help small businesses with international patent protection"), & 32 (requiring the USPTO to support "the establishment of pro bono programs designed to assist financially under-resourced independent inventors and small businesses"). *See also* Joe Matal, *A Guide to the Legislative History of the America Invents Act: Part I of II*, 21 FED. CIR. B.J. 435, 507-512.

[81] 2011 America Invents Act § 10(b)-(g), 125 Stat. at 316-18 (codified at 35 U.S.C. § 123) (authorizing satellite offices and defining a "micro entity" as an inventor with fewer than four patents and whose income did not exceed three times the median household income for the preceding calendar year).

[82] Lee, *supra* note ___, at 20-28 (including patentable subject matter, utility, and the experimental use doctrine).

[83] Lee, *supra* note __, at 51-57(citing specific examples with respect to the doctrines of novelty, priority, written description, and common law experimental use).

[84] 2011 America Invents Act § 5(e)(5). *See also* Robert Barrett et al., *Did University Patents Become the World's Most Valuable Patents Following Enactment of the America Invents Act?*, INTELL. PROP. TODAY, July 2012, at 32.

[85] 35 U.S.C. § 123 (2012); Matal, *supra* note ___, at 461 and n.165 (describing the Reid amendment entitling public universities micro-entity fee status).

though most universities can hardly be characterized as small or poor. Universities were also the intended beneficiaries of a one-year "grace period" was enacted as part of the same law.[86]

Policymakers have also paid particular attention to the class of innovators that does not make products but asserts their patents as a business model, so-called "patent trolls" or "patent assertion entities" (PAEs).[87] PAEs have distinct traits—they do not have customers and therefore are invulnerable to countersuit; they often also use economies of scale to reduce the costs of assertion.[88] In 2006, the Supreme Court changed the standard for granting patent injunctions and, in his concurrence, Justice Kennedy warned of an industry "in which firms use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees."[89] Since the rule change, non-practicing entities have generally been denied injunctive relief, rather than granted it.[90]

But not only has patent law been tailored to meet the needs and risks associated with particular classes of innovators, it has also been, on occasion, customized to specific patentholders who have sought extensions to the terms of their particular patents. The tradition of asking Congress for special treatment predates the federal patent system, and finds its root in the British tradition of issuing individual "letter patents" to requestors.[91] Requests to the U.S. Congress have been modest in number, and mostly denied.[92] However, they support the contention that the needs of particular innovators, not just the demands of particular types of innovation, have influenced the arc of patent law.

Thus far, scant empirical academic attention has been paid to innovator-specific policies outside of special interest areas like PAEs,[93] but they are in need of evaluation and assessment.

---

[86] Matal, *supra* note ___, at 479 and n.284 (citing remarks by Senators Leahy and Hatch that the grace period would protect "small and university inventors in particular").

[87] First described in Colleen V. Chien, *From Arms Race to Marketplace: The Complex Patent Ecosystem and Its Implications for the Patent System*, 62 Hastings L.J. 297,328 (2010) (defining PAEs as entities "focused on the enforcement, rather than the active development or commercialization of their patents"). *See also* Fed. Trade Comm'n, Patent Assertion Entity Activity: An FTC Study 1 (Oct. 2016), https://www.ftc.gov/system/files/documents/reports/patent-assertion-entity-activity-ftc-study/p131203_patent_assertion_entity_activity_an_ftc_study_0.pdf (defining PAEs as businesses that "acquire patents from third parties and seek to generate revenue by asserting them against alleged infringers" which "already use . . . the patented technology").

[88] Mark A. Lemley & A. Douglas Melamad, *Missing the Forest for the Trolls*, 113 Colum. L. Rev. 2117, 2129 (2013).

[89] eBay, Inc. v. MercExchange, LLC, 547 U.S. 388, ___ (2006).

[90] Christopher B. Seaman, *Permanent Injunctions in Patent Litigation After eBay: An Empirical Study*,101 Iowa L. Rev. 1949 (2016).

[91] Oren Bracha, Owning Ideas: The Intellectual Origins of American Intellectual Property, 1790-1909 191 n.8 (2016) ("Congress was presented with numerous petitions for individual grants…" et seq.).

[92] Robert P. Merges & Glenn Harlan Reynolds, *The Proper Scope of the Copyright and Patent Power*, 37 Harv. J. Legis. 45, 60, nn.59-61 (2000). Such requests have also not been limited to patent law. *See id.* (chronicling parallel private petitions in copyright and other laws).

[93] PAEs and topics including PAE patent sources have been the subject of intense academic and policy intention, much of it empirical. A May 2017 search of academic articles on Lexis Nexis and JSTOR found

Precious little is known about whether policies to encourage filing by universities or small firms, for example, are working as intended and improving participation in the patent system, or have had limited impact. Even less is known about the long-term impact of such policies on applicants or on the patent system in general. If lower prices are leading inventors to file for protection over low-quality inventions that don't ultimately mature into patents, despite the expenditure of time and money,[94] or to financial losses when the idea is not financially valuable and not worth keeping a patent over,[95] the intended beneficiaries of the policy may actually be worse, not better off.

C. Different Classes of Innovators Experience and Exploit the Patent System Differently.

The previous subpart explored the many ways in which patent law has been innovator-, not only innovation-driven. Given the long tradition in the U.S. patent system of excluding, including, and differentiating by innovator, it comes as no surprise that the lived experiences of participants in the patent system also depend on who the patentholder or participant is. Below, I discuss how differences in innovator traits have correlated with differences in innovator experiences.

Studies of the ways that jurors decide patent cases suggest that traces of past preferences can, to some degree, be found in the present patent system. In a pair of studies, then-professor, now Federal Circuit Judge Kimberly Moore, considered the outcomes of 4,000 patent cases involving domestic and foreign parties.[96] Even though the principle of national treatment requires the equal or better treatment of foreigners, in cases between domestic and foreign parties, the domestic party bested the foreign party almost two-thirds (64%) of the time.[97] The same matchup before judges resulted in a win rate much closer to even.[98] It should be noted that more recent studies have found both that foreign inventors (as opposed to foreign parties) "do just fine" in litigation as compared to domestic inventors, perhaps in part as a result of bringing only the strongest cases,[99] and also, little support for claim that the main patent trade court, the International Trade Commision, is

over 503 law references (including 350 law review articles) and 2,616 social science references (including 2,305 journal references). *See also* Edward Lee, *Patent Trolls: Moral Panics, Motions in Limine, and Patent Reform*, 19 STAN. TECH. L. REV. 113 (2015) (reporting that the top 10 U.S. newspapers had published at least 130 articles featuring the term).

[94] As discussed further in Part IV, the available data suggests that the abandonment rate among non-large entities, both before and after grant, is higher than it is for large entities.

[95] *Id.*

[96] Hon. Kimberly A. Moore, *Populism and Patents*, 82 N.Y.U. L. REV. 69 (2007) [hereinafter Moore, *Populism*]; Hon. Kimberly A. Moore, *Xenophobia in American Courts*, 97 NW. L. REV. 1497 (2003) [hereinafter Moore, *Xenophobia*].

[97] Moore, *Xenophobia*, *supra* note __, at ___.

[98] Moore, *Xenophobia.*, *supra* note __, at ___.

[99] John R. Allison et al., *Understanding the Realities of Modern Patent Litigation*, 92 TEX. L. REV. 1769, 1796 (2014).

protectionist.[100] Moore found the outcome to be even more lopsided in matchups between corporations and individuals, the latter also historically favored under patent policy as described above. When before a jury, individuals won their cases nearly three quarters of the time and corporations, only a quarter of the time, as compared to equivalent win rates before judges.[101] As with early patent laws, these jury outcomes reflect positive discrimination in favor of small and independent inventors and negative discrimination against foreigners. However, it is more likely the case that broader social norms and prejudices are influencing and shaping patent policies and outcomes rather than the other way around.[102] Here, narrative rather than empirical accounts can explain the results—stereotypes about the romanticized "heroic inventor"[103] and "intellectual property thieving foreigner"[104] continue to persist to this day, even at Presidential levels.[105]

The class, race, and gender of children with the potential to become innovators also are correlated with whether or not they participate in the patent system. As discussed in greater detail in Part III, the preference for free, white male inventors encoded in the first patent system has been followed by the dominance of this group among innovators that patent.[106]

PART II: HOW INNOVATORS

While the last Part addressed *why* innovators should be studied through the patent record. This Part addresses *how* innovators can be studied. Empirical researchers have pored over patents in search of clues about underlying innovation.[107] But the patent record is arguably

---

[100] *See* Colleen V. Chien, *Patently Protectionist? An Empirical Analysis of Patent Cases at the International Trade Commission*, 50 Wm. & Mary L. Rev. 63 (2008).
[101] *See* Moore, Populism, *supra* note ____.
[102] *See, e.g.*, Kevin M. Clermont & Theodore Eisenberg, *Xenophilia or Xenophobia in American Courts? Before and After 9/11*, 4 J. Empirical Legal Stud. 441 (2007) (discussing the relationship between perceived court bias against foreigners and their aversion to U.S. forums, which grew after 9/11, resulting in a case selection associated with elevated foreigner win rates).
[103] *See, e.g.*, Moore, *Populism*; Cotropia, *supra* note ___, at ___.
[104] *See* Colleen V. Chien, *An Empirical Analysis of Patent Cases at the International Trade Commission*, 50 Wm. & Mary L. Rev. 63, ___(tracing use of the "foreign pirate" motif throughout history to justify the ITC's Section 337 authority, despite the reality that most foreign defendants to ITC 337 cases, are competitors, not copyists).
[105] *See, e.g.*, Noah Friedman, *'They Haven't Played By The Rules': Trump Accuses China of 'Massive Theft of Intellectual Property' and Unfairly Taxing U.S. Companies*, Bus. Insider (Dec. 9, 2016, 10:07 AM), http://www.businessinsider.com/donald-trump-accuses-china-massive-theft-intellectual-property-unfair-taxing-tawian-2016-12 (add parenthetical).
[106] *See* Bell et. al., *supra* note ___.
[107] *See* Part III for an overview. This is based, for example, on the assumption that patentees put more effort into inventions that they think are more valuable. For example, researchers have looked for patents that contain indicia of high value, including a greater than average number of claims, related patents, or prior art citations. Most of the literature on valuable patents has focused on the traits of litigated patents

just as rich a source of information about the *innovators* of a patent as it is about the *innovation* of the patent. This Part begins with an overview of how patent records are constructed and the types of information they contain. Table 2A lists several classes of both innovator and innovation information available in the patent record and provides exemplar information drawn from one specific patent, U.S. Patent 8,454,073. Many of the same benefits and drawbacks of patents as measures of *innovation* extend to patents as measures of *innovators*, and this Part next discusses this topic. Finally, this Part describes data and methods for accessing each of the types of innovator data described, from the patent record directly or in combination with other off-the shelf tools. The methodology and data described leverage recent developments that have dramatically increased the availability and reduced the cost of high-quality patent big data. In support of further research and replication, we make our approaches and codebase available in Appendix A.

A. How Patents and Patent Records are Constructed[108]

A patent is a government-granted right to exclude others from the practice of an invention. However, patent rights do not arise automatically; on the contrary, patents are only granted after a representative of the inventor, typically a patent attorney or agent, prepares and submits an application to the USPTO and convinces the patent examiner that the invention deserves a patent. The application is a document that includes drawings and figures, a description of how to make and use the invention, and a set of "claims" at the end that define the scope of the rights sought.[109] The examination process generally takes several years and is called patent "prosecution." During prosecution, the representative and patent examiner typically have several rounds of legal exchanges in which the examiner will "reject" the application and the representative will adjust the application, for example, narrowing its scope. In most cases, the patent examiner will eventually grant the patent based on the refined application.

After a patent is granted, the patentee must renew it to keep it in force, approximately every 3.5 years until expiry.[110] Each step requires the patentee to pay fees to the Office, and make decisions, for example, about whether or not to keep pursuing the application and how vigorously, or whether or not to let it lapse. If the patent is later transferred, securitized, licensed, challenged, or litigated, these events may also be reflected in the patent record kept at the USPTO.[111] Along the way, much information about the inventor and the owner, as well as the underlying innovation, enters the public domain. In most cases, after eighteen months, the full patent application, as well as details about the inventors and owners of the patent, are

---

based on the assumption that litigated patents are, unlike the vast majority of patents, patents that are worth fighting over. *See, e.g.*, John Allison et al., *supra* note ___, at 465-70 (2004). For an overview of the unique characteristics of litigated patents, see, e.g., Katrin Cremers, *Determinants of Patent Litigation in Germany* 7-13 (Ctr. for European Econ. Research (ZEW), Discussion Paper No. 04-72, 2004), *available at* ftp://ftp.zew.de/pub/zew-docs/dp/dp0472.pdf (describing litigated patents as having greater numbers of citations, patent family members, and other characteristics).

[108] For an overview of the patenting process, see ROBERT P. MERGES & JOHN F. DUFFY, PATENT LAW AND POLICY 51-53 (6th ed. 2013).

[109] As discussed *infra*, the application may be accompanied by various petitions to accelerate its examination, and it may also have counterparts in other jurisdictions that undergo separate examination.

[110] Maintenance fees are due at three and a half years, seven and a half years, and eleven and a half years after the patent issues. 35 U.S.C. § 41(b) (2012).

[111] See Chien, *Predicting Patent Litigation, supra* note __, at 300-308 for detailed descriptions of these events.

published by the Patent Office.[112] Correspondence and payments between the Patent Office and the inventor's representative, are also made public as part of the "File History."

[112] 35 U.S.C. § 122 (2012) (Confidential status of applications; publication of patent applications.)`

Table 2A: Examples of Innovation and Innovator Data and Metrics[113]

Innovation Data	**Innovator Data**	
Data Category	**Data Category**	**Data from Patent 8,454,073**
Invention and Patent Dates	Inventor Name	Johannes Grandel, et al.
Abstract	Inventor Location	Friedberg (DE)
# of Claims	Inventor Citizenship	German
Patent Classification	Foreign or Domestic Inventor	Foreign
# of Words per Claim	Owner Name	Continental Automotive GmBH
Related Applications	Owner Location	Hannover (DE)
Title	Owner Size	Small
Abstract	Foreign or Domestic Owner	Foreign
Keyword search results of patent text	Government Interest	None
Office Action Rejections	Inventor Gender[114]	Male
Prior Art Considered	Inventor Ethnicity[115]	European

Information about innovators is available from three sources: patents themselves, patent records, and combinations of patent, patent record, and outside data sources. Researchers historically have used a number of techniques to manually collect these data[116] but now much of it can be gathered in bulk format. The front page of a patent lists the names and citizenship of all the inventors of a patent, as well as the owner of the patent[117] (often the company of the inventor), and the cities, states, and countries of both. Government interests in patents, typically arising from the funding of the invention, also must be recorded on the front page. Contractors

---

[113] A number of studies have considered subsets of these innovation and innovator categories of data. *See, e.g.*, Michael Risch, *The Layered Patent System*, 101 Iowa L. Rev. 1535 (combining technological and personal features of the patent).

[114] Inferred based on inventor name.

[115] Inferred based on name.

[116] *See, e.g.*, Naomi R. Lamoreaux & Kenneth L. Sokoloff, *The Decline of the Independent Inventor: A Schumpeterian Story* 4-6 (Nat'l Bureau Econ. Research, Working Paper No. 11654, 2005) (describing the use of manual collection techniques to harvest of historical patent data from 1870-1911 and 1820-1855 from sources including *Patent Gazettes,* the *Annual Reports of the Commissioner of Patents*, and the National Archives underlying a number of historical studies about the pre-modern patent system).

[117] *See* 35 U.S.C. §§ 115-16, 152 (2012) (requiring inventors to apply for patents and requiring the PTO to grant the patent to the patent owner, often the employer of the inventor, respectively).

must include statements on patents that reflect U.S. government interests that, "[t]his invention was made with government support under (identify the contract) awarded by (identify the federal agency). The government has certain rights in the invention."[118]

In contrast to the patent document, which conveys a static snapshot of the invention at the time of its issuance, the patent record reflects transactions that occur over a patent's lifetime.[119] As with all empirical work, the sourcing of innovator metrics from patent data and federation of this data with other sources must be done carefully and with full awareness of their strengths and weaknesses. Many of the factors that make patents useful but imperfect innovation metrics also apply to patents as innovator metrics. They are discussed in detail next.

B. The Strengths and Weaknesses of Patent Data for Innovators

The biggest drawback of relying on patents is the lack of correspondence between patenting and innovation, as not all innovations are eligible for patents and not all eligible inventions are patented. Many factors can prevent a novel idea from becoming the subject of a patent including the expense and difficulty of applying for or enforcing a patent, uncertainty about whether the idea is patentable, a desire to keep it secret, and the availability of alternative means of protection.[120] Firms of different sizes experience these factors differently—for example, smaller, cash-strapped firms are more likely to be more selective about what they patent than larger firms, particularly during downturns. Patentability standards have also evolved over time, in a way that unevenly impacts industries. For example, while "laws of nature, natural phenomena, and abstract ideas" have long been considered unpatentable, the pendulum that dictates the patentability of biotechnology, business method, and software innovations has swung in recent years, at times dramatically.[121] How much innovation an individual patent application represents is also highly variable, as the ratio of R&D per patent has varied considerably over time,[122] and new patent applications themselves can be derivative of existing patents as "continuations."

Patents may be granted too slowly for products in fast-moving industries like mobile apps, which can be imitated within months,[123] while the twenty-year maximum term of a patent may be too short for certain types of preventative treatments, which take decades to test.[124] An NSF survey in 2008 found that only one in eight companies that did research and development

---

[118] 35 U.S.C. § 202(c)(6) (2012); 37 C.F.R. § 401.14(f)(4) (2004).

[119] For a review of the different events that can happen after a patent is issued, and which are recorded in the patent record, see, e.g., Chien, *Predicting Patent Litigation, supra* note ___, at 300 fig.1 (describing ownership, investment, financing, citation, and enforcement events that can take place after a patent's grant).

[120] Graham et al., *supra* note ___, at 1311 & fig.4 (2009).

[121] *See* Part IV *infra*.

[122] *See, e.g.*, Colleen V. Chien, *Comparative Patent Quality*, ___ Ariz. St. Law J. ___ (2018), at 23 fig.3 (estimating a decline in the ratio of U.S. R&D per U.S. electrical equipment and computing patent in 2007 to about one fifth of its 1983 value and citing related findings).

[123] Christian Helmers et al., Innovation without Patents? Evidence from the Mobile Apps Market 11 tbl.4, 17-20 (June 11, 2014) (unpublished manuscript), *available at* https://www.dropbox.com/s/mb4yqhfulq2whzl/Helmers%20on%20apps.pdf?dl=0 (finding that 0.04% of smartphone applications in the Apple iOS store are patented, and that applications can be imitated in as few as four months by a "fast follower").

[124] Eric Budish et al., *Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials*, 105 AM. ECON. REV. 2044 (2015).

sought a patent that year.[125] Just as patents may only be filed on a fraction of the innovation that is occurring, only a fraction of patents are translated into commercialized products[126] and just as the R&D of individual patents varies, so does their market value — many novel ideas are economic losers.

In a similar way, a variety of factors impact who patents in any given year and in any given setting. When the filing firm is small, and only has a handful of patent filings, few inferences about the overall makeup of the innovators at the firm can be made based on them. Patenting at larger firms is more likely to be representative of innovation in general at the firm, but innovators whose ideas, regardless of how important, can't be or just aren't patented won't be reflected in the patent record.

Another issue with using patents as records of innovators is that they can be incomplete. Patent applications, some of which are never mature into patents, provide a more comprehensive record of patentable innovation than do grants, but the USPTO only started publishing application data in 2000, leaving over a century's worth of patent applications hidden from public inspection.[127] And this data excludes the fraction of applications that never publishes, though the USPTO has made a concerted effort to provide at least high-level statistics.[128] The USPTO record of all that happens to a patent is also incomplete due to a lack of compliance with mandatory reporting requirements[129] or lack of such requirements in the first place. For example, when a patent changes hands, registration of the change in ownership, while prudent, is not legally required. Firms fail to record these changes for a variety of reasons, some strategic and others due to administrative lapses.[130] For this reason, it can never be said with absolute certainty after a patent has issued who owns it.[131] Unlike name and location

---

[125] John E. Janowski, *Business Use of Intellectual Property Protection Documented in NSF Survey*, NAT'L SCI. FOUND. (Feb. 2012), https://www.nsf.gov/statistics/infbrief/nsf12307/.

[126] Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341 (2010).

[127] The USPTO's Office of the Chief Economist, using its privileged access to these patent applications, has supplied total applications numbers that includes applications (published and unpublished) from 1850-the present. *See* Marco et al., *The USPTO Historical Patent Data Files: Two Centuries of Innovation*, *supra* note __, at 30 fig.4 (showing varying differences between the number of applications filed and granted over time)

[128] The USPTO's Office of Chief Economist has also sought to include this information in its aggregate disclosures. *See id.*

[129] For example, according to 35 U.S.C. § 290, trial courts must let the PTO know when a patent is litigated, as well as when a judgment on such litigation is issued. 35 U.S.C. § 290 (2012). Yet as many as 35% of litigations are never reported, according to Robin Feldman et al., *The AIA 500 Expanded: The Effects of Patent Monetization Entities*, 17 UCLA J.L. & TECH. 1, 65 (2013). Likewise, an analysis of government interest statements among academic biomedical patents from 1980 to 2007 period found systemic under-disclosure. Arti K. Rai & Bhaven N. Sampat, *Accountability in Patenting of Federally Funded Research*, 30 NATURE BIOTECH. 953, 954-955 (2012) (finding a reporting rate of sixty to ninety percent among known government-funded patents in the same period).

[130] *See* Changes to Require Identification of Attributable Owner, 79 Fed. Reg. 1405 (proposed Jan. 24, 2014) (to be codified at 37 C.F.R. Part I) (describing the USPTO's proposal for making recordation of attributable owners mandatory for patent records. After opposition, particularly after the costs of compliance had surfaced, the USPTO shelved its proposal. Developments described at https://www.uspto.gov/patent/initiatives/attributable-ownership.

[131] Re-assignment data is messy, because it is recorded alongside many other types of changes to the status of a patent. Alan C. Marco et al., *The USPTO Patent Assignment Dataset: Descriptions and Analysis* (U.S. Pat. & Trademark Off., Working Paper No. 2015-2, 2015).

inventor information, which are collected from every patent when granted,[132] these additional fields of data are not uniformly available.

While none of these factors by itself precludes the careful use of patents as measures as innovators, they do underscore the risk of mischief and misuse of patent data, and the importance of selecting appropriate metrics and drawing supported inferences. In this study we apply several "best practices" to reduce the risk that observed trends reflect the behavior of innovators rather than behavior in patent lawyers.[133]

First, we primarily couch our findings in relative, not just absolute terms - e.g. regarding the rise in the share of Asian inventors, or the relatively higher representation of female inventors in certain classes and jurisdictions relative to others. Second, we limit truncation effects by relying primarily on the metric of granted patents, rather than applications or citations. For the reasons described above, focusing on patent grants rather than applications enables a longer period of observation and avoids the temporal truncation bias associated with the lag between applications, publications, and grants. However, if there are particular types of patents that issue faster than others (e.g., patents from a certain state, or of a certain technology area) this will slightly skew the later data points. In addition, though as described later, citation counts have certain advantages, not relying on them reduces truncation biases. To minimize the risk of changes in regional or national policy and industry or firm-level patenting to skew the results, we study cumulative patenting over all firms for several decades, minimizing the bias associated with any one change; and make note of major known shifts. Finally, we shy away from making causal inferences about innovation based on patent records alone.

These precautions enable the exploitation of several advantages of the patent record relative to other modern innovation records. The patent record is long- reliable records are available starting from 1836.[134] The patent record is accessible - after all, the word patent derives from the word "patere" which means "to lay open,"[135] and one of the two enumerated duties of the United States Patent and Trademark Office is to disseminate information about patents.[136] Unlike academic publishing, another "open innovation" platform, patents don't risk being trapped behind paywalls and dead links.[137] In addition, despite the limitations described earlier, there are few if any comparable source of information about innovative activity at the firm, inventor, and invention levels across sectors. Thus, while highly imperfect, patent records are uniquely accessible, longitudinal, and connectable to demographic data; that is to say, as

[132] The first patent—Patent X1, to Samuel Hopkins of *Philadelphia* on August 4, 1790—was one of only three granted that year. http://invention.si.edu/united-states-patent-certificate-manuscript-granted-samuel-hopkins-1790 (emphasis added). It was signed by President George Washington and endorsed by Thomas Jefferson. *Id.*

[133] Much in accordance with the "Checklist for Analyses" set forth by Josh Lerner and Amit Senu, *The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond* 6, 19 tbl.10 (Nat'l Bureau of Econ. Research, Working Paper No. 24053, Nov. 2017).

[134] The record from 1790-1836 is spottier, due to a Patent Office fire in 1836.

[135] *Patent*, ONLINE ETYMOLOGY DICTIONARY, https://perma.cc/3P9S-X2G9 (last visited Feb. 5, 2017); *see also* SUBCOMM. ON PATENTS, TRADEMARKS, & COPYRIGHTS, S. COMM. ON THE JUDICIARY, 85TH CONG., AN ECONOMIC REVIEW OF THE PATENT SYSTEM 1 (Comm. Print 1958) (prepared by Fritz Machlup) (describing the noun "patent" as the customary abbreviation of "open letter" or "letters patent," a literal translation of the Latin *litterae patentes*).

[136] 35 U.S.C. § 2(a)(2) (2012).

[137] For a description of open innovation systems, see Kevin J. Boudreau & Karim R. Lakhani, *"Open" Disclosure of Innovations, Incentives and Follow-on Reuse: Theory on Processes of Cumulative Innovation and a Field Experiment in Computational Biology*, 44 RES. POL'Y 4 (2015).

Zvi Griliches once did, they are an oasis in the "desert of [innovation and innovator] data."[138]

C. Data Sources and Methods

[**note to reader: you most likely will want to skip this section as it will likely go into an appendix later and also misses information about the WIPO data]

This study draws on public patent records associated with U.S. utility patents issued from around 1870 to around 2016,[139] with a focus on the last four decades.[140] Each decade is represented by the sixth year of the decade (e.g., 1986, 1996). For each year, except in the case of 2016 due to truncation effects, two months were selected, representing one-sixth of the patents from the year.[141] This study cross-references other studies and official aggregate statistics provided by the USPTO, including by state,[142] by nationality (U.S. or foreign)[143] and independent invention. This study also draws upon international patent filings from 2005 to the present obtained from WIPO.

1. Inventors, Owners, Patents, Companies, Organizations

This study considered several metrics in connection with each studied patent. For the studied years from 1986-2016, the name of each inventor was extracted and his or her city, state, and country from each patent record, and used as the basis of additional analyses. This study reports on the traits of inventors associated with patents issued the studied years, counting an inventor each time she appeared in the record. Because we counted each inventor individually, rather than assigning co-inventors fractional values in proportion to the number of inventors, patents with multiple inventors are, in a sense, "overrepresented" in the data. This study also relied on the USPTO's record of patent ownership at the time of issuance, although ownership can change at any time in the lifetime of a patent application or patent.[144]

To compensate for the wide variation in patent value, we observed for one analysis not only the quantity of inventing, through patent counts, but the quality of inventing, relying, as

---

138 Zvi Griliches, *Patent Statistics as Economic Indicators: A Survey*, 28 J. ECON. LIT. 1661, 1661 (1990).

139 In particular, this study primarily relied on PATENTS VIEW, http://www.patentsview.org/ available through BigQuery, bigquery.cloud.google.com/table/patents-public-data:patents.publications_latest (last visited Feb. 5, 2018) (for records from 1976 to present) and INNOGRAPHY, https://www.innography.com/ (last visited Feb. 5, 2017) (for pre-1976 records), supplemented by hand coding using AMAZON MECHANICAL TURK, https://www.mturk.com/mturk/welcome (last visited Feb. 5, 2017) for missing fields.

140 For details, see Appendix, Table 1A.

141 March and September, except in the case of 2016, when we relied on the only available data we hadm from March 1-22 of that year, due to lag in the updating of PatentsView in BigQuery, to be updated when the data is refreshed.

142 *See, e.g.*, PATENT TECH. MONITORING TEAM (PTMT), U.S. PATENT & TRADEMARK OFFICE (USPTO), *Patent Counts by Country, State, and Year - Utility Patents (December 2015)*, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cst_utl.htm (last visited Feb. 5, 2017); PTMT, USPTO, *Independent Inventors by State by Year* (Feb. 2016), https://www.uspto.gov/web/offices/ac/ido/oeip/taf/inv_all.pdf.

143 *See, e.g.*, PTMT, USPTO, *U.S. Patent Statistics Chart, Calendar Years 1963 - 2015*, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm (last visited Nov. 5, 2017).

144 In the vast majority of cases, each patent is owned by a single owner at any given time.

others have,[145] on the number of times a patent is cited by another, subsequent patent (its "forward citation count") as a proxy for patent value.[146] This study counted the number of forward citations in the five years after grant,[147] for 1976-2016 patents, sampling where there were size constraints.[148] Like with patent counts, there are significant limitations and hazards of using patent citations as measures of patent quality.[149] However, we limit these risks by calculating citations at a single point of time, limiting temporal effects, and using citations to compare the relative performance of states, which are not known per se to experience citation behavior differently. In the analysis of the geography of inventing, this study presents results that are normalized by each state's or district's population, based on U.S. Census records, resulting in patent and cite densities. [add description of WIPO]

2. The Characteristics of Inventors

While many studies have relied on patents as measure of innovation, much less attention has been paid to patents as measures of innovators. The richness of the patent record supports the tracking of innovators by profile, firm, firm-size, geography, time, industry, and invention setting based on information provided directly by the applicant or coded by the USPTO. But while the patenting process does not capture the sex and ethnicity of each inventor, it does list the name and location of each inventor and owner, and based on this information, several demographic traits can be estimated. Kerr has pioneered the use of commercial marketing techniques and databases to identify the ethnicity of individual inventors, assigning name and location pairs to one of nine likely ethnicities.[150] Cook has constructed a database of black inventors from 1963 to 2006, and together with Kongcharoen, has developed ways to identify likely female and African-American inventors, based in part on U.S. Census data.[151] Each approach has its strength and limitations. For example, the Kerr database, though capable of making fine distinctions between Asian names,[152] leaves out many ethnicities and

---

[145] *See* discussion, *supra* notes ____.
[146] As discussed *infra* in Part I.
[147] Previous analysis has found that patent filings and citations are correlated at the state level. Philippe Aghion et al., *Innovation and Top Income Inequality* 19, 47 tbl.1 (Nat'l Bureau Econ. Research, Working Paper No. 21247, 2015).
[148] Due to file size limitations, for patents issued in the years 1996, 2006, and 2011, I estimated yearly forward citation counts based on per patent average forward citation counts for patents issued during the first three months of the year (January through March). A comparison of the twenty states with top citation counts of that list, generated based on patents issued during 1986, to the same list of states based on the patents issued during the first three months of 1986 indicated a high level of correspondence, with only a 1- state deviation.
[149] Described by Lerner and Senu, *supra* note ___, at 6, 19 (describing the variable ways in which patents have come to be cited, and both the truncation and temporal biases associated with forward citations).
[150] Chinese, English, European, Hispanic, Indian, Japanese, Korean, Russian, and Vietnamese; For a review of this work and the techniques Kerr has developed, see William R. Kerr, *U.S. High-Skilled Immigration, Innovation and Entrepreneurship: Empirical Approaches and Evidence* (World Intell. Prop. Org., Econ. Research Working Paper No. 16, 2014), and Foley & Kerr, *supra* note ___.
[151] Described in Cook & Kongcharoen, *supra* note __.
[152] For example, distinguishing between Korean and Chinese names.

does not distinctly identify, for example, African names. The Cook database of African-American inventors, though one of the most comprehensive ones available, only extends through 2006 and does not identify other ethnicities or races. Both require access to privately- developed and maintained databases.

This study relied upon publicly available classifiers that use name and location information to make predictions about an individual's binary gender[153] and ethnicity. For example, Kim and Lois are first names are typically associated with females, and Peter and Steve, with males. However, not always - the classifier Genderize.io assigns 90% and 94% probabilities, respectively, to the odds "Kim" and "Lois" are female [154] while "Peter" and "Stevie" are associated with a 100% and 63% probability, respectively, of being male.[155] In addition, context matters. While Genderize.io gives the English name "Andrea" a 98% chance of being a female; the Italian name "Andrea" from Italy has a 79% chance of being male.[156] In contrast to gender predictors, which rely primarily on first names, ethnicity classifiers depend on both first and last names. The distributions vary; e.g. Martha Gonzales has a 93% chance of being Hispanic, whereas Betty Gonzales has about a 54% chance of being Hispanic and a 40% chance of being White, using the classifier we relied on in this study.[157]

A classifier's ability to make an accurate prediction depends on the quality and quantity of both the data on which the classifier is trained and the data to be classified, as well as their relationship. If a trainer is classified on a population that predominantly includes men, for example, but the population to be profiled is drawn from society at large, the results are likely to be skewed in favor of men. Further, omissions or ambiguities in input data cannot be reliably be profiled. Thanks to the USPTO's extensive normalization efforts, the patent names relied upon were substantially free of foreign or unparseable text, however, even following the application of extensive cleaning protocols,[158] certain names could not be resolved — for example, some because of low input quality (e.g., with first names including "â„«strand", or the single letter "T"), and others because they are unique and not covered by the databases (e.g., "Hinsdale" and "Wirt").[159] Both missing and low quality data reduced the share of inventors to whom likely gender and race qualities could be assigned, to 99% and 97% of inventors by race and gender, respectively.[160]

---

153 The sex of a person generally finds its basis in his or her biological, anatomical features, whereas gender is generally understood as social and cultural in nature.

154 DETERMINE THE GENDER OF A FIRST NAME, https://genderize.io/ (last visited Mar. 22, 2018).

155 *Id.*; NATURAL POTATO MAGNET, GENDERIZE, https://www.npmjs.com/package/genderize (last visited Mar. 22, 2018).

156 *Id.*

157 NAMEPRISM, http://www.name-prism.com/ (last visited Mar. 22, 2018). Disambiguating by nationality raises additional challenges, for example, as certain names, like Chen can be both Chinese and Korean. I am thankful to Bill Kerr for sharing this insight with me.

158 For example, that decoded unicode characters into a normalized form. See the Appendix for a link to the cleaning protocols we used.

159 For example, BenoÃ®t Belier, Love â„«strand, Juan (n)}ano (author's research).

160 For counts, see Appendix A.

Keeping these cautions in mind, we took several steps to increase the accuracy of our approach while also ensuring its scalability and reusability. First, we used two classifiers for each trait, both to ensure that our results were not sensitive to our classifier choice, and to increase coverage. For gender data, we primarily used Genderize.io, a general purpose gender classifier independently recognized as having a higher performing algorithm than other commercial methods.[161] Based on a first name and country input, the classifier returns the probability of the individual being male or female, or a finding of "no match." Because our analysis focused (although not exclusively) on U.S.-based inventors, for a second gender classifier, we used the U.S. Social Security Baby Names Database, which uses public records o include the names of babies born in the United States from 1880 to the present and which has been relied upon by other researchers.[162] The agreement rate between the two sources was 98.3% and the coverage from the Genderize.io classifier, across time exceeded 99.9%, so we report exclusively on the Genderize.io classifications.

For race and ethnicity data, this study used NamePrism, a nationality classifier developed based on a dataset of 57M names, also determined to have the highest accuracy score among competing classifiers, with coverage of 39 groups representing over 90% of the world population.[163] The NamePrism algorithm accepts name information (first and last), and returns a probability distribution among five possible ethnic groupings, including Black, Asian-Pacific Islander, White, Hispanic, and American Indian and Alaska Native (AIAN). For a second ethnicity classifier we referred to upon the Census 2000 names table, which includes surnames occurring 100 or more times in the 2000 Census and their distribution among the following groupings: White, Black, Asian-Pacific Islander, American Indian and Alaska Native, and two races.[164] Their agreement rate[165] was 97.88%. Though the coverage for each ethnic classifier was around 80%, when we combined the two classifiers, as others have done,[166] our coverage rate improved to 96.43%.[167] We exclude AIAN and two-race results from the distributions reported herein.

---

[161] Fariba Karimi et al., *Inferring Gender from Names on the Web: A Comparative Evaluation of Gender Detection Methods*, 25 Int'l Conf. Companion World Wide Web Proc. 53 (2016) (finding Genderize.io to perform best among individual detection methods). Further, the test data on which the detection methods were benchmarked consisted of a test set of academic researchers, a reasonably close proxy of inventors.

[162] Karimi et al., *supra* note ___.

[163] The underlying algorithm is described in Junting Ye et al., *Nationality Classification Using Name Embeddings*, 2017 Conf. Proc. __.

[164] Available and described at *Frequently Occurring Surnames from the Census 2000*, Census.gov (Sept. 15, 2004), https://www.census.gov/topics/population/genealogy/data/2000_surnames.html.

[165] Based on a "highest probability" approach as described *infra*.

[166] Described, e.g., in Rebecca Knowles, Josh Carroll, & Mark Dredze, *Demographer: Extremely Simple Name Demographics*, 2016 EMNLP Workshop on Nat. Language Processing & Computational Soc. Sci. 108.

[167] The coverage rate across time ranged between 96.2% and 97% For details, see Appendix A.

There are at least two approaches to estimating gender and nationality distributions across a population: to aggregate probabilities across the population ("aggregate probability estimation"), or, for each record, to assign a value based on the highest probability option ("highest probability estimation"), and then sum up the totals.[168] To minimize the risk of making our findings dependant on the approach we chose, we tried both. We found both estimation approaches to yield almost identical results as to inventor gender distribution, and chose to apply highest probability estimation throughout the paper.[169] As to inventor ethnic distribution, the results were mixed. We observed that distributions of Hispanic, Asian, and to a degree White names did not depend too much on the estimation technique, but that the share of Black inventors varied dramatically, with a cumulative estimation approach yielding a much larger share.[170] As a result, we disclose in each view the approach we display and note and report on both sets of values where the difference is significant (>10%).

Several additional factors limit the accuracy and completeness of our results and should be kept in mind when interpreting them. First, the classifiers themselves are imperfect. For example, though Nameprism reports the highest cumulative accuracy or "F1" score among publicly available classifiers, it is still only 0.795, out of a highest score of 1.[171] In addition, classifier accuracy varies substantially by ethnic group — European and Asian names can be identified with greater accuracy than Greater African names,[172] contributing in part to the sensitivity described above of the results to the estimation approach. This weakness of available databases with respect to their identification of Black names is an impediment to tracking black inventors and deserves further attention. None of the classifiers, as far as we are aware, quantifies the characteristics of its unmatched names but if they are skewed in any particular direction, this could introduce bias into the reported rates.

3. The Characteristics of Patent Owners

The characteristics of patent owners can provide further insights about innovators. Since at least 1976, for example, the USPTO has placed each patent into one of nine major

---

[168] Take, for example, a dataset that included 10 people, each with a 40% probability of being male. Choosing the highest probability would result in all individuals being profiled as female, for a total of 10 females. Aggregating probabilities, on the other hand, would result in a 40% male distribution among the total, or 6 females and 4 males. See Ye, *supra* note ___ for further exploration of these comparative approaches.

[169] A less-than-1% difference in any year or technology sector category. Detailed results available upon request.

[170] This we suspect, based on an interview with Tamyra Walker of YesWeCode, is an artifact of the legacy of slavery and forced adoption, in many cases, of master names by former slaves. As such, the difference between the "highest probability" and "cumulative probability" shares of black inventors increased from less than 1% to 5% or 6%, and the shares of white inventors declined by a commensurate or greater amount while Asian shares were within a 0.5% difference and the Hispanic shares were within a 1.3% difference.

[171] Ye, *supra* note __, at 1.

[172] *Id.* at 2 tbl.2. (reporting lower classifier F1 scores in association with Greater African names as compared to Greater European and Asian names regardless of the classifier.)

categories, owned by: a U.S. company or corporation, foreign company or corporation, U.S. individual, foreign individual, U.S. government, foreign government, country government, or state government (U.S.), or none (unassigned).[173] Because of the particular interest in the evolution of independent inventors, this study used a technique similar to ones used by other scholars[174] to identify independently-invented patents to supplement the official statistics. To assess independent inventorship: 1) this study assumed that unassigned patents were independently invented, 2) in the case of assigned patents, the last name of the inventor was compared to the owner of the patent. If a patent was assigned, but the name of the assignee contained the name of one of the inventors, it was also coded as independent-inventor owned. For example, assuming that the inventor "David Kellogg" owned, at least in part, the company "Kellogg Switchboard and Supply Company." In the earlier, pre-1976 data, this sort of assignment to what appears to be the inventor's own business was more common. Applying this method to a century of patents, the trend using the USPTO's codings was similar to the trends observed using this study's approach, showing a decline from 12% to 6% from 2002 to 2015 (USPTO codings)[175] versus a decline of 9.5-14% to 6% from 1996 and 2006, respectively, to 2016 using my approach.[176]

The category of "corporation" or "company" includes the smallest LLC and the largest public corporation. To qualify as "small," a patent owner (or licesee) must be either an individual inventor, university of other qualifying nonprofit, or a business entity with fewer than 500 employees,[177] a number that is tied to the Small Business Administration's definition of a small business concern. Microentities, which are entitled to the steepest discounts, must meet the small entity requirements and, in addition, be employed by a university or other qualifying institute of higher education, or have a gross income less than three times the median household income in the U.S. and not have been named on four prior applications within a specified time frame.[178] Whether or not a patentee has paid a reduced fee can be discerned from the file on a patent by patent basis. For the purposes of this paper, the Office of Chief Economist of the USPTO generously made available the entity codings associated with a 1.5% random sample of patents from 2000 to the present.[179]

4. The Geography of Innovators

When a patent is granted, the Patent Office prints the city, state, and country of the inventors and the owner of a record on the patent.[180] This study estimated the location of each

---

[173] The USPTO also assigns codes to patents that are partially owned, however, the share of patents with partial ownership codes was so small that we did not include them in our ownership analysis.
[174] For a description of a similar methodology applied to early records, see Lamoreaux & Sokoloff, *supra* note __, at 21.
[175] Using the following sources of official USPTO data: PTMT, USPTO, *Patent Counts by Country, State, and Year*, *supra* note __; PTMT, USPTO, *Independent Inventors by State by Year*, *supra* note ___.
[176] *See* Figure ___ in Part III.
[177] 13 C.F.R. § 121.802 (2011); 37 C.F.R. § 1.27(a)–(b) (2011).
[178] 37 C.F.R. § 1.29(a)–(b) (2012).
[179] For a total of about 50,000 patents, of which 85% patents were coded. *See* Appendix Table A2.
[180] This practice dates back to at least 1976 and is currently recorded on the front page of the patent. While "front page data" is as of the time of grant and is not updated following issuance of the patent, changes in ownership and inventor location are often recorded with the Patent Office and can be discerned by examining assignment and related records.

patent by coding the locations of the owners. It uses normalized U.S. city and state assignee locations for the U.S. assigned patents, and the location of the first-named inventor for unassigned patents to approximate the inventor location associated with the patent. Bulk data about invention locations is only available as of 1976, so Mechanical Turk's coders inspected and manually hand-coded the locations associated with 0.05% of the patents from the decades from 1896 to 1976. The results were the same whether the NBER or WIPO categories were used.[181]

One problem with using classification-based codes is that the assignment of a patent to a particular code is not always done consistently and is sometimes the target of strategic behavior by applicants. In addition, although patents are often assigned to multiple classification codes, particularly as research becomes more combinatorial, for the sake of simplicity, this analysis relies on the first code associated with each patent. In order to represent city data graphically, it was mapped to zip codes and counties,[182] which were used as the basis for the heat maps shown. To track the extent to which patenting tracks population density relied upon aggregate data provided by the USPTO that tracks patent counts by "core based statistical area" (CBSA),[183] geographic areas of relatively higher density that may cut across city, town, or county lines.[184] These counts were further divided into Metropolitan Statistical Areas (MSAs), "areas with at least one urbanized area of 50,000 or more inhabitants" and Micropolitan Statistical Areas, defined as having "at least one urban cluster of at least 10,000 but less than 50,000 population."[185]

5. Technology Sector

To support innovator comparisons, this study classified patents and inventors by technology sector. To do so, patent classification codes were leveraged and assigned to each patent by the Patent Office for the purpose of assignment to an examiner with the relevant expertise applying "prior art" references. From 1790 to 2014, the USPTO assigned U.S. Patent Classification (USPC) codes to all incoming patent applications.[186] In January 2015, the USPC was replaced by the Cooperative Patent Classification (CPC), a joint effort with the European

---

[181] Author's own analysis (showing substantial increases in the electrical and computers and communication (NBER) and electrical engineering (WIPO) shares and decreases in the drugs and medical (NBER) and mechanical, chemical, and instrument (WIPO) shares from the 1970s to the present).

[182] Using *U.S. Census Concordance*, U.S. CENSUS BUREAU, http://www2.census.gov/geo/docs/maps-data/data/rel/zcta_county_rel_10.txt (last visited Mar. 22, 2018).

[183] PTMT, USPTO, *Calendar Year Patent Statistics (January 1 to December 31): General Patent Statistics Reports Available for Viewing*, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports_cbsa.htm (last visited Feb. 5, 2017).

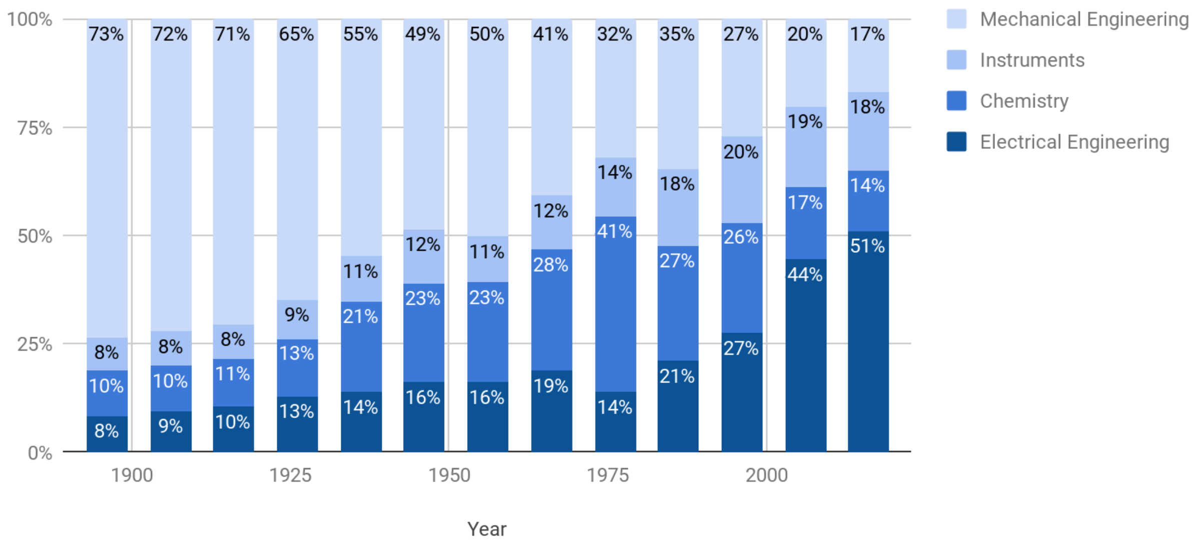
[184] For example, in California, the "Redding," "Modesto," and "San Francisco-Oakland-Hayward" MSAs.

[185] PTMT, USPTO, *Patenting in U.S. Metropolitan and Micropolitan Areas: Breakout By Organization*, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cbsa_asg/explan_cbsa_asg.htm (last viewed Feb. 5, 2017); U.S. Census Bureau, *Metropolitan and Micropolitan Statistical Areas of the U.S. and Puerto Rico* (Feb. 2013), *available at* https://upload.wikimedia.org/wikipedia/commons/1/16/Metropolitan_and_Micropolitan_Statistical_Areas_%28CBSAs%29_of_the_United_States_and_Puerto_Rico%2C_Feb_2013.gif.

[186] For a timeline of the history of USPC and CPC codes, see *Making the Switch from USPC to CPC: What Attorneys and Searches Need to Know*, TECH. & PAT. RES. (Oct. 2015), http://www.tprinternational.com/making-the-switch-from-uspc-to-cpc/.

Patent Office (EPO).[187] As the codes have evolved over time, the USPTO has reclassified patents into the newest set of categories, and in 2014, the USPTO assigned CPC codes to patent documents from 1836 to December 2014. There have been at least two efforts to aggregate the hundreds of patent classification codes, which are used to facilitate patent examination, into industry classifications that can be used for research and policy analysis. The NBER Patent Citations Data file classified patents based on USPC code into six main categories and 36 sub-categories.[188] WIPO has promulgated another scheme based on CPC classes that include five main categories and 35 subcategories.[189] The present analysis relies on WIPO technology categories, as applied to the first CPC code of each patent. However, the reported observed shifts in technology shares to the present were the same regardless of methodology.[190] As Fig. 2A shows, there has been a dramatic shift from the dominance of mechanical engineering patents at the turn of the century, with almost 75% of the classified patents, and the continued emergence of "manufacturing" fields (including mechanical, chemical, and instruments) through 1976, to, over the last four decades, the growth in the share of digital (electrical engineering) patents from 14% in 1976 to 51% in 2016. (Fig. 2A)[191]

FIG 2A: Grant Shares by Technology Area (1896-2016)



This Part has outlined sources and methods for gaining insights about innovators that patent, and the hazards of doing so. The next Part applies these methods to patent records with a focus on the last four decades, to explore and establish basic demographic facts about those who participate in patented innovation.

---

187 *Id.*

188 Hall et al., *NBER Patent Citations*, *supra* note __, at 13.

189 ULRICH SCHMOCH, CONCEPT OF A TECHNOLOGY CLASSIFICATION FOR COUNTRY COMPARISONS: FINAL REPORT TO THE WORLD INTELLECTUAL PROPERTY ORGANISATION (WIPO) 9 tbl.2 (2008).

190 Author's analysis.

191 [Author's data: https://docs.google.com/spreadsheets/d/1zOj7-4pVAaMWHnU6WLbthX7q8lpmzIOCtH-w75oixBk/edit?zx=jii1q6tpl4t6&usp=docs_web]

PART III: WHO, WHERE, AND IN WHAT SETTINGS ARE INNOVATORS THAT PATENT?

The question of who is participating in innovation, and in what setting has long been of interest. As recounted in Section I, the early patent system welcomed certain classes of innovators while excluding others. Though most legal barriers to patenting were dismantled over a century ago, the extent to which patenting specifically and innovation generally reflects inclusive and democratic ideals remains open and relevant. For example, Congress recently charged the USPTO with "'establish[ing] methods for studying the diversity of patent applicants, including those applicants who are minorities, women, or veterans.'.[192] Carrying out this mandate, the Patent Office attempted to connect patent records to Census records, in order to develop aggregate demographic statistics. However, the match rate was too low (64.3%) to be useful.[193] The Patent Office then explored collecting demographic data directly from applicants. But when stakeholders objected to mandatory reporting, the USPTO shelved the idea.[194] The methods described in Part II provide ways to circumvent these obstacles to confirm patterns of patenting at scale and reveal novel, combinatorial insights about the evolving profile of innovators that patent.

A. Who Is Inventing?

[**a note of caution to the reader: this section needs a lot of work and rework in view of the AI case study in Part IV - feel free to skip]

1. The Gender Diversity of Inventors

The gender gap in science and engineering education and workforce participation is well-documented. Women make up about half of the college-educated U.S. workforce, but only 24% of the STEM workforce.[195] While women earn nearly three out of ten STEM degrees,[196] each field admittedly with a different likelihood of patenting, they are named on only one in ten U.S.-origin patents.[197] Patent applications by women are less likely to mature into patents and patents by women are less likely to be commercialized (as measured by assignment rate),

---

[192] Leahy-Smith America Invents Act § 29, Pub. L. No. 112-29,125 Stat. 284, 339 (2011). For an overview, see USPTO, Study and Report on the Implementation of the Leahy-Smith America Invents Act 48 (2015), [hereinafter *USPTO Report*]. This provision was introduced in order to, in the words of its author Rep. Gwen Moore, overcome the gap in information that "prevents us from fully understanding the nature and scope of the underrepresentation of minority communities in intellectual property." 157 Cong. Rec. H4,484 (daily ed. June 23, 2011) (statement of Rep. Moore).

[193] *USPTO Report*, *supra* note __, at 48.

[194] *Id.*

[195] ITIF, *supra* note ___, at 8 n.8.

[196] Jessica Milli et al., *Equity in Innovation: Women Inventors and Patents*, Inst. for Women's Pol'y Res. 9 fig.5 (Nov. 29, 2016), https://iwpr.org/publications/equity-in-innovation-women-inventors-and-patents/ [hereinafter Milli et al., *EII*].

[197]Jennifer Hunt et al., *Why Don't Women Patent?* 1 (Nat'l Bureau Econ. Research, Working Paper No. 17888, 2012); *accord infra* Figure _ (documenting a 9-10% female inventor rate).

though they are at least or more likely to be cited.[198] The gap appears to be smaller in academic settings than in corporate or government environments, perhaps because of the less hierarchical nature of universities.[199] And the gap is narrower than what it once was. From 1790 to 1860, only 77 patents were invented by women, compared to 4,773 patents to men in 1860 *alone*, translating into a female patenting share of approximately 0.00018%.[200]

A closer look at rates of female patenting, including by sector and country, provides a broader context for understanding some of these trends. Based on applying the methods described in Part II to data on inventors from 1986 to 2016, we found, consistent with others, a significant increase in the participation of women in inventing over the last three decades.[201] On average, women comprised around 5% of all U.S. inventors in 1986 but more than double that, 12%, in 2016. Growth has been uneven, however. In 1986, across sectors, female inventors comprised about 3-6.5% of U.S. inventors. But by 2015, the female share of U.S. inventors had grown to ~19% in "chemical industries,"[202] which includes a wide swath of biopharmaceutical, medical device, and related sectors, but only ~6% and~ 9% of "mechanical engineering" and "instruments" patents, respectively. (Appendix, Fig. 3A) More granular inspection of the 2015 record shows that of 35 industry sub-sectors, biotechnology and then organic chemistry, fields that are perceived to have flatter and more flexible firms,[203] have the largest shares of female inventors, respectively 22% and 20%.[204] The findings of strong industry effects and overall growth, as also documented by others,[205] confirm the use of the off-the shelf, quantitative approaches described here.

While gender inclusion in inventing in any particular industry setting has many inputs, making it difficult to point to any one particular driver of change as the most important, related

---

[198] ITIF, *supra* note __, at __ fig.8 (67% of women patent applications are granted patents, as opposed to 73% of applications by males); *id.* at __ fig.9 (30% of patents with a female primary inventor were unassigned, as compared to assignment rates of 20% or less among other types of patents); *id.* at 15 (citing the finding of Cook & Kongcharoen, *supra*, that median citations for U.S. women inventors are at least as high as overall citations across technology areas).

[199] Cassidy R. Sugimoto et al., *The Academic Advantage: Gender Disparities in Patenting* 8, PLoS ONE (May 27, 2015).

[200] Khan, *supra* note __, at 162-64.

[201] For a recent summary of the literature that uses a variety of different metrics, see Milli et al., *EII*, *supra* note ___, at 7-8 (reporting that the share of patents having at least one female inventor increased from 3.4% in 1977 to 18.8% in 2010 and that the share of patents where the primary inventor was a woman increased from 2% to 8% over the same time period).

[202] *See* Schmoch, *supra* note __, at 9 tbl.2.

[203] Gema Martinez et al., *Identifying the Gender of PCT Inventors*, World Intellectual Property Organization (WIPO) Economic Research Paper No. 33 (Nov. 2016), at 3-4.

[204] Author's analysis.

[205] *See* Milli et al., *EII*, *supra* note__, at 10 fig.6 (finding chemistry, organic compound, and drug-related classes to have the highest shares of female patentees); Paola Giuri et al., *Everything You Always Wanted to Know About Inventors (But Never Asked): Evidence from the PatVal-EU Survey* 7 (Lab of Econs. & Mgmt. (LEM), Sant'Anna Sch. of Advanced Studies, LEM Working Paper Series, Feb. 2006), http://www.lem.sssup.it/WPLem/files/2005-20.pdf (reporting, based on European Patent Office records from 1995-1997 that the gender disparity was the least pronounced in chemicals and pharmaceuticals (7.4% female) and the most pronounced in mechanical engineering (1.1% female)).

work suggests that pipeline effects are also significant.[206] In 2010, women in the United States captured over 50% of biological/biomedical degrees but less than 15% of engineering degrees, and a declining share of computer science degrees,[207] corresponding with inventing shares. However the pipeline is leaky - women's inventoship representation is still, at ~19% of inventors in "chemical" industries and 6.6-11.3% of engineering inventors, only a fraction of the rate of their academic representation. Setting also matters. In a study of 12,000 *academic* scientists, Ding and her colleagues found that academic women patented at approximately 40 percent of the rate of men with equivalent credentials and networks, a significant gap, but one much less than the gap between male and female innovators in corporate settings.[208] Further probing the differences between commercialized and non-commercialized patents, Hunt found that only 7% of the gap in commercialized patents could be explained by women's underrepresentation in science and engineering.[209]

2. The Racial Diversity of Inventors

Just as inventing has become more gender-diverse, it has also become more racially diverse. When we estimated shares of inventors by race during this time using the methods of Part II, we confirmed that the share of non-white U.S. inventors had grown substantially from 1986 to 2016,[210] from around 7-16% to 26-32%. (Appendix, Fig. 3B) Again, the gains have been distributed unevenly. As others have also found,[211] our methods showed that the share of Asian inventors as a percentage of all U.S.-based inventors has grown dramatically, from 6% in 1986 to ~ 23-24% in 2016. The shares of inventors based in the U.S. that are Hispanic have also grown, but remain far below their shares of the general population. In 1986, the share of Hispanic inventors as a portion of all U.S. inventors was, using the methods described in this paper, estimated to be about 1-2.4% (based on highest and cumulative probability approaches, respectively). (Appendix, Fig. 3B) By 2016, this share had grown to an estimated 2.5-3.5% of inventors, as compared to a 17.8% population share within the United States.[212] As to black inventor shares, the differences in outcomes based on highest and cumulative probability

---

[206] Though it should be noted that the divergence in paths that potential innovators take starts much earlier than college, and encompasses a variety of social and cultural conditions, including "biases and barriers that begin in preschool and persist through the workplace." Kapor Center, *supra* note __, at 7.
[207] Milli et al., *EII*, *supra* note ___ , at 9 fig.5; *accord Fraction of Bachelor's Degrees in STEM Disciplines Earned by Women*, AM. PHYSICS SOC'Y (last visited Aug. 20, 2017), http://www.aps.org/programs/education/statistics/womenstem.cfm.
[208] Waverly W. Ding et al., *Gender Differences in Patenting in the Academic Life Sciences*, 313 SCI. 665 (2006). *Id.*
[209] Hunt, *Why Don't Women Patent?*, *supra* note ___, at 2. *See also* Catherine Ashcraft & Anthony Breitzman, *Who Invents IT?: Women's Participation in Information Technology Patenting, 2012 Update*, NAT'L CTR. FOR WOMEN & INFO. TECH. 3 (2012) (finding that nine percent of IT patents have at least one female inventor, a share that has increased over time but remains far below parity).
[210] From 16% in 1986 to 33% in 2016 based on a cumulative probability estimation approach.
[211] *See Kerr*, *supra* note ___ (finding that the share of Indian and Chinese inventors has increased dramatically over the period 1975 to 2004, from under 2% to 6% and 9%, respectively).
[212] *FFF: Hispanic Heritage Month 2016*, U.S. CENSUS: FACTS FOR FEATURES (Oct. 12, 2016), https://www.census.gov/newsroom/facts-for-features/2016/cb16-ff16.html.

estimation approaches are too great to provide useful point estimates, but according to any metric, are just a fraction of their share in the general population of 13% in 2016.[213]

Racial disparities in inventing, like gender disparities, are stark. In the case of black inventors, de facto exclusion and non-participation have replaced legal exclusion. The low participation of Latinos and Latinas in innovation relative to their representation in the general population also suggests that much talent lies untapped.[214] Asians, in contrast, have dramatically grown their share, and the share of white inventors, as a result, has declined, from over 80% to closer to 70%, depending on the estimation method used. (Appendix, Fig. 3B). But has the decline in the *share* of white inventors also correlated with a decline in the *number* of white inventors? When the absolute rather than relative values are considered, it becomes clear that *number* of white inventors has grown substantially, however the *growth* of Asian inventors has outpaced the growth of white inventors. (Appendix, Fig. 3C). For that matter, the growth of all non-white groups has outpaced that of white inventors, from much smaller starting shares in 1986. (Appendix, Fig. 3C)

B. Who Owns Patents, In What Settings?

Closely related to the question of who is patenting is the question of who owns patents and in what settings patenting is happening. As discussed in Part I, from the start, the U.S. patent system has included features intended to level the playing field between corporate and independent inventors. In addition to the fee policies described in that Part, over time, U.S. law has included several other unique features designed to favor small and independent inventors, facially and in practice.[215] But how does the participation of small and independent inventors in patenting through the decades match their position and role in shaping patent law?

---

[213] *QuickFacts: United States*, U.S. Census Bureau, https://www.census.gov/quickfacts/fact/table/US/PST045216 (last visited Mar. 23, 2018).

[214] *Id.*

[215] The development of many of these features has, in turn, been influenced by the independent inventor lobby. For example, in support of the idea that under resourced individuals should have the same rights as corporations to pursue patents, the United States patent system has, until recently, featured a "first to invent" approach to determining patent rights. This approach is seen as fairer to independent inventors than the "first to file" approach adopted by the rest of the world because it rewards having the idea first, not necessarily having the patent application on file first. Cotropia, *supra* note ___, at 66.

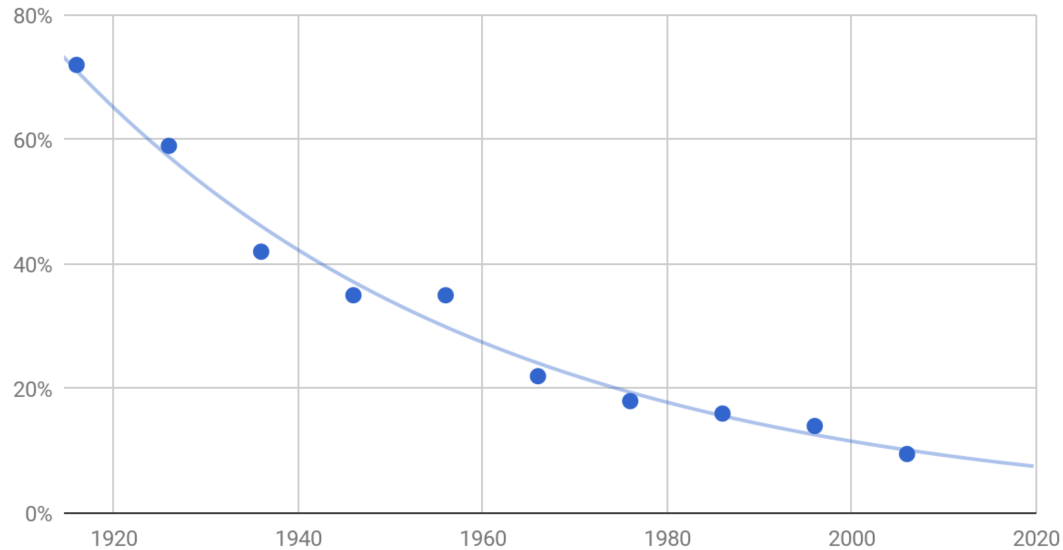
Although the 2011 America Invents Act transitioned the U.S. to a first to file system, in order to harmonize the U.S. to international law, the U.S. retained a unique "grace" period specifically in response to demands by and in order to accommodate the needs of independent inventors and universities. Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011); Matal, *supra* note ____, at 457-62, 479 n.282 (citing statement of Sen. Orrin Hatch, who describes the provision as one that "will benefit independent and university inventors in particular.")

A decade or so before the America Invents Act, Congress was also poised to require mandatory publication of patent applications, consistent with the practice of other countries. But the independent inventor lobby argued vociferously against the idea, concerned that pre-grant publication would make it easier for large corporations and others to steal their ideas. *See, e.g.*, Mark Janis, *Patent Abolitionism*, 17 Berkeley Tech. L.J. 899, 919-920 (2002); *see also* An Open Letter to the U.S. Senate, Eagle Forum, http://www.eagleforum.org/patent/nobel_letter.html (last visited Sept. 1, 2016) [https://perma.cc/P6A5-RSPA] (letter by twenty-four Nobel laureates criticizing the change and claiming that it would be "very damaging to American small inventors and thereby discourage the flow of new inventions").

1. Independent Invention Through the Decades

FIG. 3F shows the share of patents developed independently over the last century or so. While in 1906, the overwhelming majority of patents were independently invented,[216] in 2016, only about 6% of patents were unassigned or assigned to an individual at the time of patenting. (FIG. 3D) The most dramatic declines occurred well before 1976: from 1906-1976, the average

Fig. 3F: Independent Inventor Shares Over Time



decline per decade in the share of independently invented patents was 9%; from 1976-2016 the average decline was 3%. This steep declines of the earlier era accompanied the broader transitions experienced by the United States from, as Mokyr has described, a system in which inventors were "the hero[es] of the modern age....the main actors who brought on the Industrial Revolution" to a system driven by corporate research and development.[217]

Independent inventorship has continued to decline, from about 20% in 1976 to about 5% in 2016, which provides important context when considering the input of the independent inventor lobby. This share only represents raw patent counts, without correcting for their relative importance, but related research has found that independent inventor patents are cited and

---

The final rule included an exception allowing those only filing their applications in the U.S. to opt out of publication, benefiting independent filers. 35 U.S.C. §§ 122(b), 154(b) (2006).

[216] *Accord* Naomi R. Lamoreaux et al., *The Reorganization of Inventive Activity in the United States during the Early Twentieth Century* (Nat'l Bureau of Econ. Research, Working Paper No. 15440, 2009) (showing that 71% of patents were not assigned in 1890–1891).

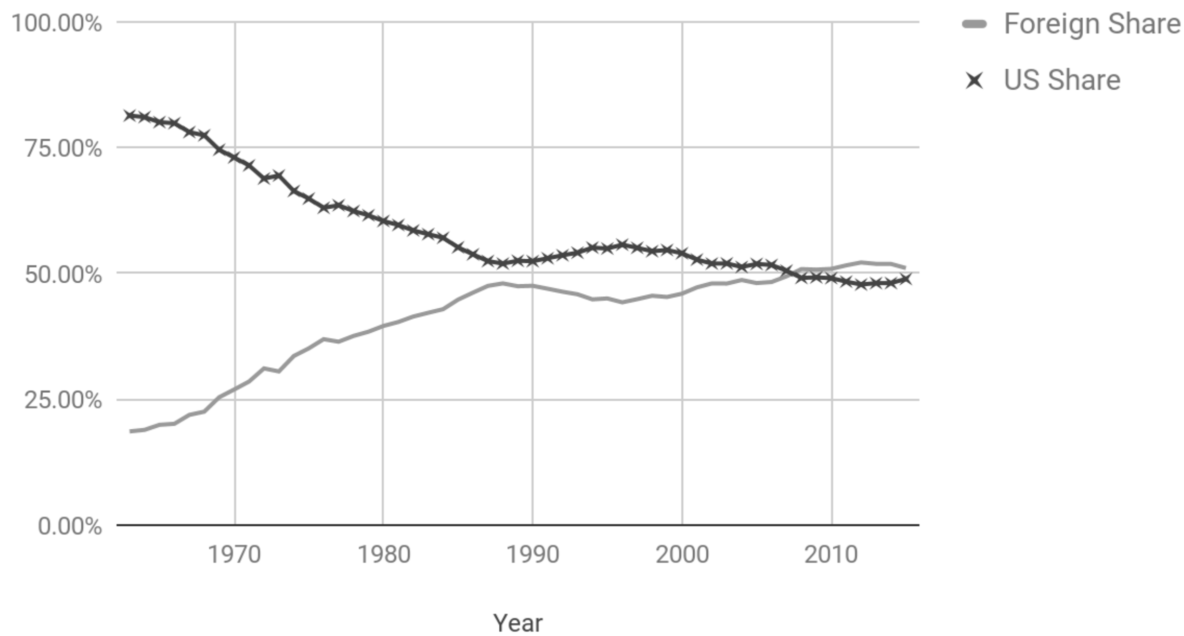
[217] Merritt Roe Smith et al., *Historical Perspectives on Invention & Creativity* 22-23, 2003 Lemelson-MIT Program Workshop Proc., http://web.mit.edu/monicaru/Public/old%20stuff/For%20Dava/Grad%20Library.Data/PDF/history-3289136129/history.pdf.

renewed less.[218] If independent inventor shares have declined, what shares have increased? To answer that question, the next subpart considers not only the extent to which inventions are arising in corporate or independent settings but the places and traits of patent owners.

2. The Settings of Invention

FIG. 3E shows shares of patent ownership based on assignee type over the past four decades. During this period, U.S. government shares of patents, and U.S. company shares have changed modestly in terms of absolute percentages: from around 3% to closer to 0-1%, and from 45% to 46%, respectively. However, while independent inventor shares have declined, from 20% to 6%, foreign company shares, which encompass the activities of foreign governments and individuals, have increased from 30% to 50%. (Appendix, Fig. 3G) have gradually increased over time. Driven by strong patenting by Asian firms, foreigners have filed for more U.S. patents than Americans since 2009.[219] (Fig.3H)

FIG 3F: US and Foreign Shares of Utility Patents



The decline in independent inventorship and increase in foreign ownership of U.S. patents over the past four decades is striking in several ways. First, it supports populist accounts of economic transition according to which foreign interests are taking the place of

---

[218] Jim E. Bessen, *The Value of U.S. Patents by Owner and Patent Characteristics* 26-27 (Boston Univ. Sch. of Law, Working Paper Series, Law & Econ., Working Paper No. 06-46, 2006) (finding based on citation and renewal data that small and individual inventor patents are worth less than are patents from large firms, and discussing studies that have reached similar conclusions).

[219] *Utility Patent Application & Grant Count (2 Visuals)*, USPTO Open Data Portal, https://developer.uspto.gov/visualization/utility-patent-application-grant-count-2-visuals (last visited Aug. 20, 2017).

independent, domestic stakeholders. At the same time, however, it underscores the extent to which protectionist measures within the patent system that are framed as measures to defend U.S. innovators against foreign interests are limited in their ability to do so. These limits are illustrated by the example of the ITC, whose 337 authority empowers it to defend domestic industries against infringing imports. Because of the offshore manufacturing of products by American companies and the diversification of patent ownership, U.S. companies are targeted as much as if not more than foreign firms, the data suggests.[220] Finally, the transition in inventing setting and location supports a reframing of the American patent system as one that protects not only American innovators but to an increasing degree, the interests of foreign innovators as well. The American intellectual property system is often positioned as a bulwark for domestic innovators against foreign pirates. However, if current trends continue, the U.S. patent system will more often than not provide shields to foreign rather than domestic patent owners.

C. Where Patenting?

1. The Domestic Geography of Patenting

Putting aside the question of the international geography of patenting, the domestic distribution of patenting is also of outstanding policy interest. As discussed earlier, the design of the early patent system included a number of features intended to encourage broad-based innovation. Though the data is sparse, several data points suggest these efforts met some success. In her analysis of a sample of “great inventor” patents from 1790-1930, Khan found that nearly half of the inventors in the studied sample, like Thomas Edison, had little or no formal schooling.[221] In her study of British and American innovations at world fairs between 1851 and 1915, Moser found that there was a large disparity between urban and rural patenting rates in Britain, but no systematic difference in patenting rates between urban and rural areas in the U.S.[222]

The importance of place to innovation is widely acknowledged.[223] Over several decades, scholars have confirmed the importance of spatial and geographic dimensions, as well as trends like urbanization, localization, and diversity, to innovation.[224] To understand how the locus of

---

[220] *The International Trade Commission and Patent Disputes: Hearing Before the H. Comm. on the Judiciary*, 112th Cong. 22 (2012) (statement of Colleen V. Chien, Professor, Santa Clara University School of Law) (showing that, from the 18-month period from Jan. 2011 to June 2012, 209 domestic defendants were named in ITC 337 actions brought by patent-assertion entities vs. 123 foreign defendants).

[221] Khan, *supra* note ___, at 183-90.

[222] Petra Moser*, Innovation without Patents: Evidence from World's Fairs*, 55 J.L. & ECON. 43 (2012).

[223] Maryann P. Feldman & Dieter F. Kogler, *Stylized Facts in the Geography of Innovation in* 1 HANDBOOK OF ECONOMICS OF TECHNICAL CHANGE 381 (Bronwyn H. Hall & Nathan Rosenberg eds., Elsevier B.V. 2010). *See also* Richard Florida et al., *The City as Innovation Machine* (Martin Prosperity Inst., Working Paper No. 2016-MPIWP-002, 2016). Geography matters not only to where invention takes place, but how it spills over. Despite reductions in the cost of disseminating information, local patents are more likely to be cited by an inventor than similar patents from elsewhere. Florida et al., *The City as Innovation Machine*, *supra* note __, at 6.

[224] Feldman & Kogler, *supra* note __, at 1.

inventing has shifted over time and build on previous literature,[225] I looked at the cities and states of U.S. patents over the decades. In 1873 the states with the highest patents per capita were inland states North Dakota and Montana. In 1976, the states with the highest grant densities were Pennsylvania, New Jersey and California, all states on the coast. In that year, five of the top five[226] and eight of the top 10 states[227] were coastal states. (Appendix, Table A5). By 2015, nine out of ten were coastal states.[228] (Appendix, Table A5)

What has made certain geographies more innovative than others? Universities have been important drivers of inventing – in 2015, 8 of the 10 states with the highest per capita patent grant densities – California, North Carolina, Pennsylvania, Virginia, Texas, Georgia, New Jersey, Maryland – were among the states with the most universities, colleges, and institutes of higher learning.[229] (Appendix, Table A4) While most of these states are among the largest, others are not – Connecticut and Iowa have some of the highest grants per capita, buoyed by their institutes of higher learning. And thus, while the numbers of patents granted to universities are relatively low,[230] the importance of colleges and universities to local innovation is much, much higher.[231]

The relevance of urbanization to innovation is also reflected in the patent record. In contrast to the Moser's finding reported earlier, that in the early 1900s that there was not a noticeable difference in World Fair innovation between U.S. urban and rural areas,[232] now the difference could not be much starker. In 2015, 96% of domestic patents named as their first inventor someone from a high- population density metropolitan statistical area (MSA), reflecting a consistent year over year rise since the year 2000.[233] Less than 5% of 2015 patents had a lead inventor from a non-metropolitan area.[234]

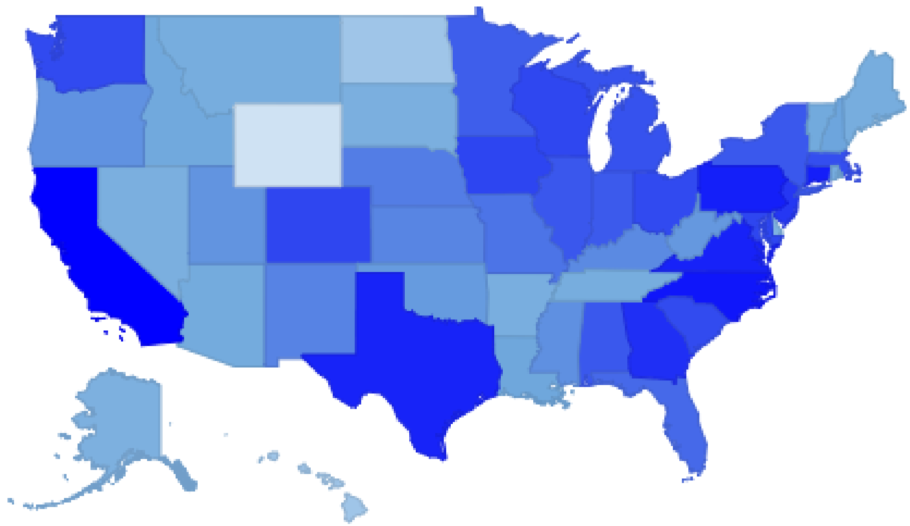
By itself then, the finding that innovation is increasingly clustering in coastal, urban, and university locations, might not be too surprising. To a large extent, urbanization in patenting mirrors the broader demographic shift of individuals to metropolitan areas, which by 2010 were home to 83% of the U.S. population.[235] In addition, for decades, scholars have observed that industries tend to agglomerate in certain locations in order to gain efficiencies in production and support specialization.[236] Innovative, educated people that comprise what Florida calls "the creative class," thrive in areas where there are a diversity of people, a culture of openness,

---

[225] *See, e.g.*, Khan and Sokoloff, *supra* note ___.
[226] California, New Jersey, Connecticut, Pennsylvania, and Texas.
[227] Maryland and North Carolina.
[228] California, North Carolina, Connecticut, Pennsylvania, Virginia, Texas, Georgia, and Maryland. The only inland state in the top 10 is Iowa.
[229] Appendix, Table A4.
[230] 1-2% in 2015. Author's analysis based on PATENTSVIEW.
[231] Although not determinative, Feldman and Kogler, *supra* note ___, finds that, based on reviewing two decades of literature, local universities are necessary but not sufficient for innovation.
[232] Moser, *supra* note __.
[233] Author's calculation, based on data provided in PTMT, USPTO, *Calendar Year Patent Statistics (January 1 to December 31)*, *supra* note __.
[234] *Id.*
[235] Paul Mackun & Steven Wilson, *Population Distribution and Change: 2000 to 2010*, U.S. CENSUS: 2010 CENSUS BRIEF (Mar. 2011), https://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf.
[236] This literature is reviewed in, e.g., Glenn Ellison et al., *What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns*, 100 AM. ECON. REV. 1195 (2010).

tolerance, and advanced technology.[237] Globalization has made it easier for them to find their ways to places like Austin, New York City, and Raleigh-Durham.[238]

Fig 3G: 2015 Annual Patent Grants Per 10K Population (Log)



But what happens when innovators leave for the coast, not from another country, but from an inland, less innovative location? The talent, energy, and potential economic development associated with the innovator leave as well. The extent to which the clustering of innovation is contributing to what I call a "domestic brain drain"[239] of individuals from inland and rural areas to coastal and metropolitan areas is a subject I leave for later analyses.

2. **Conclusion**

This Part has used patent data to explore broad patterns in the gender, race, settings, and locations of inventors and patent owners. Over time, the data shows that inventors and owners of patents have become more diverse, though the pace and extent of diversification have been moderate and uneven. Inventors have also become more concentrated in coastal states, and within those coastal states, concentrated in coastal metropolitan areas. Continuing a decades-long trend, the rate of independent invention has declined. The trends and results described here lay the foundation for future work. They demonstrate how the sources and methods described in Part III can be used to determine various demographic facts and patterns about innovators over time. They illustrate the various dimensions along which *innovators*, separate and apart from innovation, can be observed, for example, by large or small entity inventor groups, by gender, by race, by owner or inventor country, state, or city, or any

---

[237] Richard Florida, *Cities and the Creative Class*, 2 CITY & COMMUNITY 3 (2003).
[238] *Id.* at 9.
[239] Subject of work in progress, "The Domestic Brain Drain."

combination thereof. Finally, as explored in the next Part, they support *innovator*- and not just *innovation*- focused research agendas.

PART IV: DIRECTIONS IN INNOVATOR RESEARCH - 3 CASE STUDIES

How might the data and methods described in the previous Parts support new directions in research about *innovators* that patent, not just patented *innovation*? The last part explored their usefulness in uncovering broad trends in innovators, but the same approaches can be applied to specific research and policy questions. This Part demonstrates the relevance of *innovator* centered approaches to a wide variety of contexts by reference to three topics, one drawn from patent policy, one from innovation policy, and one from corporate human resources policy. Adopting an intersectional approach, the first mini-case study considers the race and gender of innovators to reveal a surprising source of Asian advantage in innovation - its relatively higher levels of gender diversity. The second mini-case study considers the disparate trajectory of diagnostics innovators of different profiles following changes in the law of patentable subject matter. The third mini-case study evaluates the effect of the introduction of a new set of policies introduced by the America Invents Act of 2012 intended to increase participation of the smallest innovators in the patent system. Collectively, they illustrate the various ways in which research focused on *innovators* can enrich and improve policymaking and decision-making about *innovation.*

A. The Surprising Asian Advantage in (Artificial Intelligence) Innovation

The question of which country has the edge when it comes to artificial intelligence is of great political, economic, and private sector interest. Numerous outlets have reported on differences in compare government investment, startups, and firm investment. Yet one of the most critical and yet hardest to track inputs for innovation are the people creating artificial intelligence innovations. The advantages of patent data over other sources of information perhaps is clearest when tracking them [describe...]

Much attention has been paid to the emphasis that the Chinese government and private sector have paid to growing the country's artificial intelligence capabilities. Less attention has been paid however to how they are accomplishing this growth. Applying the tools to the datasets described in Part II reveals a surprising source of comparative advantage enjoyed by China and some of its Asian counterparts in artificial intelligence: its greater gender diversity....

Figure 1: Female Inventor Shares Among International AI Patent Applications, by Nationality (2005-2017)

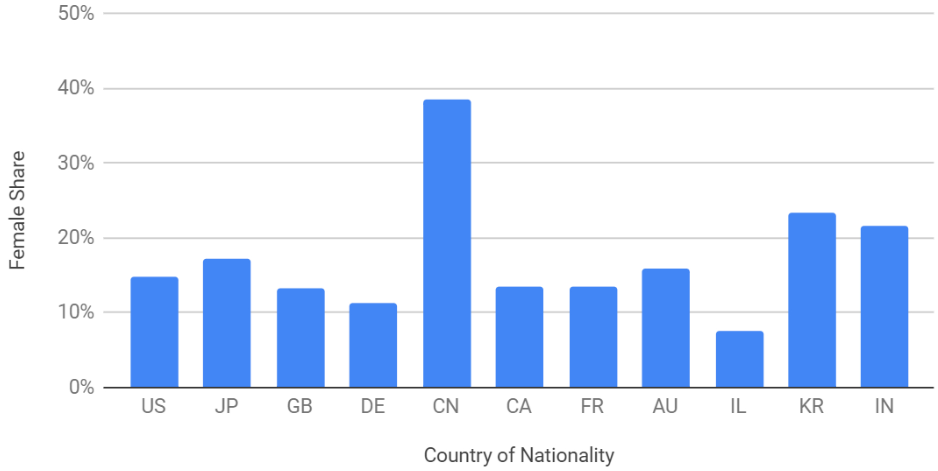


Figure 2: Female Inventor Shares Among International AI Patent Applications, by Residence (2005-2017)

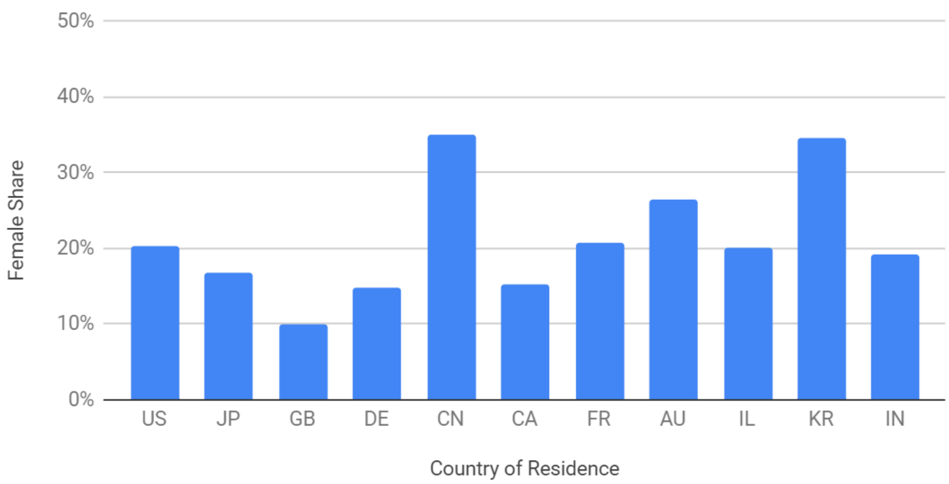


Figure 3: Female Inventor Shares Among US AI Patents (2006-present)

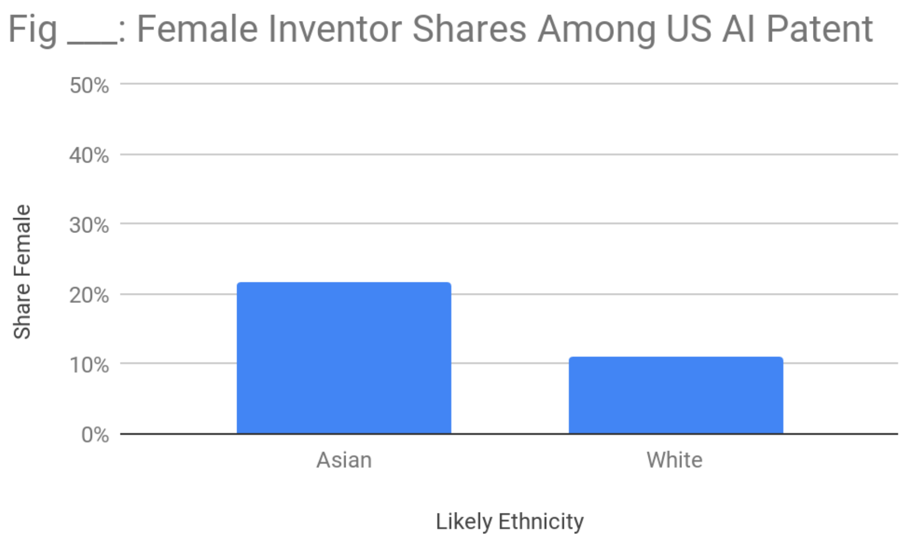
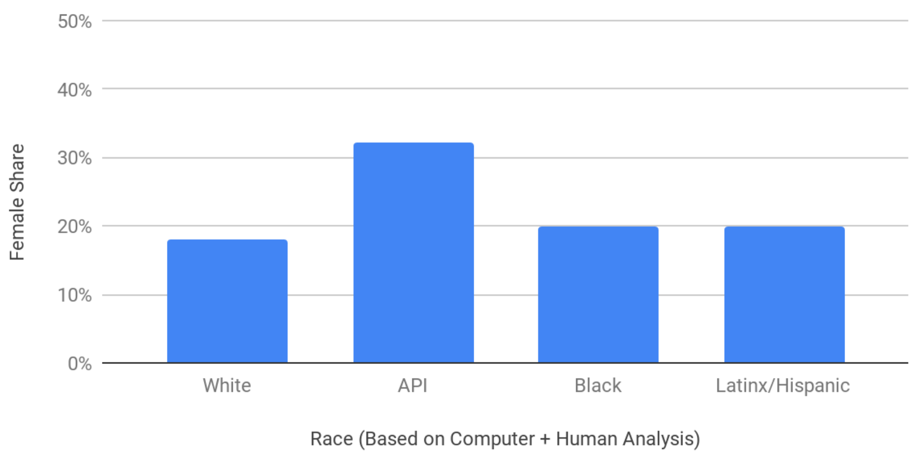


Figure 4: Female Shares Among AI Workers on LinkedIn, by Likely Ethnicity (composite based on 30 companies)



B. How Have Changes in Innovation Policy Impacted Innovation in Medical Diagnostics?

While the previous mini-case study considered the impact of innovator-specific policies, similar methods can be used to evaluate the impacts of general policy changes on particular groups of innovators. One such change took place in 2012 when the Supreme Court decided *Mayo Collaborative Services v. Prometheus Laboratories, Inc.*,[240] the upshot of which was to limit the patent eligibility of certain diagnostic innovations, for example, for detecting genetic

[240] Mayo v. Prometheus, 566 U.S. 66 (2012).

abnormalities associated with breast cancer.[241] There has been widespread concern from interest groups, the Federal Circuit, and academics that this case and related developments have harmed incentives to invest in "precision" rather than "one-size-fits-all" approaches to medicine.[242] However, it is also the case that the price of diagnostics tests has declined with greater competition, and that reducing the number of patents on diagnostics increases freedom to operate and combine tests into single kit products.[243] Citing a desire to stabilize incentives for diagnostics innovation, several groups have advanced legislative proposals to change the law.[244]

One way to gauge the impact on innovation of the changing landscape of protection is to look at the extent to which the rate of new patent filings has changed since they were issued. If less investment is going into diagnostics innovation, one would expect to see less innovation and fewer new patent filings, all things being equal. Even if innovators are still innovating, they still might reduce their fillings if patents are difficult to get, instead opting for trade secret protection.

Yet, an analysis of the rate of patent filings before and after the decision by Rai and Chien found a surprising result: that there had been no measurable decline in published patent applications, or granted patents, following the decision.[245] Rather than decreasing in number, total patent applications and grants increased steadily from 2012 to 2016.[246] How can the concerns described above that incentives for diagnostics innovation have been harmed be

---

[241] For an overview of the decisions, see Rachel Sachs, *Innovation Law and Policy: Preserving the Future of Personalized Medicine*, 49 U.C. Davis L. Rev. 1991 (2016), at 1907-1913.

[242] *See, e.g.*, Ariosa Diagnostics, Inc. v. Sequenom, Inc., 809 F.3d 1282, 1286 (Fed. Cir. 2015) (order denying for request for en banc hearing) (Dyk, J., concurring) (stating that the changes "might discourage development and disclosure of new diagnostic and therapeutic methods in the life science"); AIPLA Legislative Proposal on Patent Eligibility (May 12, 2017), http://www.aipla.org/resources2/reports/2017AIPLADirect/Documents/AIPLA%20Report%20on%20101%20Reform-5-19-17-Errata.pdf (reporting that legal uncertainty created by the changes has "weakened the U.S. patent system and discouraged investments in . . . life-saving diagnostic tools"); Sachs, *supra* note __, at 1881 (stating that together with other changes in the regulatory landscape of diagnostics, the system has "gone too far in disincentivizing desperately needed innovation in diagnostic technologies").

[243] Amelia Rhinehart *Myriad Lessons Learned,* 5 U.C. Irvine L. Rev. 1147, 1177 (2015) (describing, before *Myriad*, "claims of reduced innovation due to increased prices for tests using patented genes," and following *Myriad*, that "diagnostic companies escape the heavy royalty burden that existed and should be able to offer tests that provide a wide range of sequencing for patient").

[244] *See* AIPLA, *supra* note___; *see also* Warren Woessner, *IPO, AIPLA and ABA IP Section Propose Legislative Fixes for Section 101*, Patent4Life (May 2017) http://www.patents4life.com/2017/05/ipo-aipla-and-aba-ip-section-propose-legislative-fixes-for-section-101/ (summarizing legislative proposals by the Intellectual Property Owner's Association and the ABA IP Section proposing, consistent with the AIPLA proposal, to reduce patentable subject matter-based barriers to patent law).

[245] Colleen Chien & Arti Rai, *Jan 2017 Submission to USPTO* (2017), https://www.uspto.gov/sites/default/files/documents/RT2%20Comments%20Colleen%20Chien%20and%20Arti%20Rai.pdf.

[246] Colleen Chien, *Precision Medicine in a Post-*Bilski *and* Mayo *World* 12-13 (Apr. 20, 2017) (unpublished presentation) *available at* https://docs.google.com/presentation/d/1SY5LuzJL_QWxbSDXNATIFp9u-ROyGIYjXsxu-tkkdVs/edit#slide=id.g1996da7144_123_108. *But see id.* at 14 (finding that a control group of patent applications appears to have grown more)

squared with the relatively robust pattern of filing? A closer look at the types of *innovators* filing for protection provides one answer. According to Rai and Chien's analysis, out of about 3,000 diagnostics patent application filings in 2016, more than two-thirds were filed by Universities, nonprofits, and firms with less than $1M in revenue. Such innovators do not primarily depend on patent exclusivities in the market to sustain their research but instead, rely on grants to support publication, whether or not the research is patentable.[247] What about patent filings by companies outside of the nonprofit and academic sectors? Based on Rai and Chien's analysis, a much smaller slice (less than 10%) of overall diagnostics patenting is performed by small, for-profit companies, with revenues of $1M-$100M.[248] In contrast to the steadily increasing patent filings and grants of the university and nonprofit sector, this sector has experienced uneven growth in diagnostics filings and patents in the years following the Mayo decision, with net flat to negative growth from 2012 through 2016.[249] While small in a share of overall diagnostics patents, this sector is of critical importance because of its role commercializing inventions. The data supports that their filings have been less than robust university and nonprofit innovators following the decision.

Segmenting new filings and grants by innovator type in this way enables a more nuanced understanding of the ways in which changes in the law are - and are not - changing diagnostics innovation and patenting. Based on overall patent filings, only diagnostics innovation has not slowed following the *Prometheus* decision. However, a look at who is innovating tells a different story: patenting by small, for-profit companies, an important sector for bringing advances to the market for patient use, has been uneven, while patenting by universities and nonprofits has continued to grow. This suggests that the broad patent reforms described earlier may be motivated primarily by the needs of this particular segment of innovators.

C. Have Efforts to Support Small and Independent Inventors Worked?

As described in Parts I and III, the U.S. patent system has for years included features to encourage the participation of small and independent inventors in the patent system. However, to my knowledge, systematic attempts to assess the impact of particular, small inventor-friendly policies like "first to invent" and reduced fees have largely been lacking.[250] Starting in 2000, firms with less than 500 employees, independent inventors, and nonprofits have been eligible to get "small entity" fee discounts of 50% off of these fees.[251] Starting in early 2013, the smallest filers, as well as universities, have been eligible for even deeper "micro entity" discounts of 75%

---

[247] For a description of the differences between university and commercial patenting, see Lee, *supra* note __, at 5.
[248] Chien & Rai, *supra* note ___, at 1-2 figs.1-3.
[249] *Id.* at 2 fig.3.
[250] Studies that predict, rather than attempt to isolate, the impact of policy changes are more common. On the transition to first-to-invent, see David Abrams & R. Polk Wagner, *Poisoning the Next Apple? The America Invents Act and Individual Inventors*, at Abstract (2013) (unpublished manuscript), *available at* http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1377&context=faculty_scholarship (finding a transition to first-to-file to "likely to result in a reduced share of patents granted to individual inventors).
[251] For cites, see Part II *supra*.

off.[252] These discounts are meant to remove barriers to patenting for cost-sensitive applicants,[253] and, based on the assumption that participation in patenting is somewhat elastic, to stimulate greater filings and participation in the patent system by small and resource-constrained innovators.

Fig. 4A: Small Entity Share of US Patent Filings

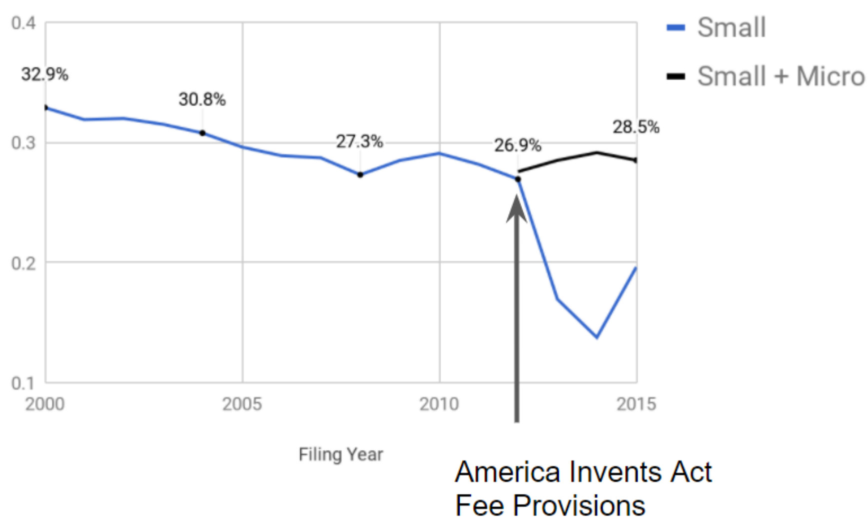


FIG. 4A shows shares of published patent filings by entity size from 2000 to 2015 based on a 1.5% sample of filings. As it shows, in 2000, nearly a quarter of all patent applications were filed by small entities. While little data on entity filings is available from before this time, it appears that the 2000 share reflects a decline from 1991, when 34% of filers were small entities.[254] It is unclear how much of this filing has been made possible by the fees discount,[255]

The share of small entity filers grew slightly, then contracted from 2003 to 2012, when about 5% of filings were by small entities. (FIG. 4A) How did the introduction of the micro-entity discount in 2013 impact the participation of small entities? One can look at changes in the rates of small and micro-entity filings for an estimate. Because the statute requires micro-entity filers to also meet the criteria required of small entities to discern the net impact of the new fee tier, this study looked at the total "non-large" share over time. As would be expected, the share of small entities declined in the first few years of the micro-entity program, as entities eligible for the deeper discount took advantage of these lower costs. However, micro-entity filings more than made up the difference with the total share of non-large (small and micro-entity) filings

---

[252] The fee change is described at *New Fees and Micro Entity Status Take Effect March 19,* Inventor's Eye (Feb. 2013), https://www.uspto.gov/custom-page/inventors-eye-advice.

[253] Several accounts have identified the high cost of patenting as the reason for not patenting. See, e.g., Milli et al., *EII*, *supra* note __, at 18-20 for an overview.

[254] H.R. Rep. No. 102-382, at 13 (1991), *reprinted in* 1991 U.S.C.C.A.N. 1320, 1328.

[255] For at least the reason that PTO fees are generally only a small fraction of the out-of-pocket costs paid by patent applicants, see, e.g., IP Watchdog, *The Cost Of Obtaining a Patent In the U.S.* (2015), http://www.ipwatchdog.com/2015/04/04/the-cost-of-obtaining-a-patent-in-the-us/id=56485/ (calculating the total cost associated with filing a patent application for a small entity as ~$12,000 to ~$22,000, less than $1000 of which are patent office fees).

growing from 5% in 2012 to 10% in 2015. (FIG. 4A) This finding is striking for several reasons. First, it represents an *increase* in the share of non-large entities that have filed for patents starting in 2013, in contrast with the steady, year over year *decrease* in non-large filings over the previous decade, from 27% in 2003 (and 34% in 1991)[256] to 5% in 2012. (FIG. 4A) Second, the amount of the share increase from 2012 to 2015 is considerable, from 5%[257] to 10%,[258] and statistically significant.[259] Third, the reversal of direction in small entity share correlates with the fee change. Because this observation is based on a small sample, this study leaves validation and further exploration of the causal effects of the fee change to future analyses when more extensive records are available.

However, it would be worth probing the extent to which the fee-change had its intended effect of causing greater participation in the system by small and micro entities. Using entity type data as described in Part II, the relative participation and increases in filings by various small and micro-entities - independent inventors, universities, and small businesses - can also be determined. For example, based on such data, according to Congressional testimony, independent entities represented 72% of small entity filers in 1991.[260] But it is unknown how these trends have varied, and whether the fee tier introduced in 2013 caused more filings from universities or small startups.

However, while greater shares of patent applications are one thing, but if they are not accompanied by greater patenting this would imply that small entities are starting the patenting process more, but completing it less. Related research suggests that the share of small entities whose patents are actually issued is smaller than the share of small entities applying for patents.[261] To the extent low fees are encouraging the ultimately unsuccessful pursuit of patents, this could be an adverse, not positive consequence.[262] This initial work demonstrates the power and promise of using the approaches described in this paper to granularly evaluate such innovator-driven policy questions.

CONCLUSION

To date, scholarship about the patent system, like the patent system, has primarily been in the service of *innovation*. This Article has made the case for reconceptualizing the patent system to understand and promote *innovators* that patent, rather than just patented *innovation*. As the Parts above demonstrate, patent data is accessible, personal, and longitudinal. It is also as yet untapped for the purposes of informing policies at the juncture of technology and human capital development. The numerous American patent policies that are innovator- not only innovation-, driven can also benefit from a shift in how we think of the patent system. Finally, recent developments in big data sources and methods have created new, powerful ways to

---

256 *Id.*
257 Out of 2,534 coded entities.
258 Out of 920 coded entities (full data was not yet available for 2015).
259 Chi-square value = 2.26572E-11**.**
260 H.R. REP. No. 102-382, at 13 (1991), *reprinted in* 1991 U.S.C.C.A.N. 1320, 1328.
261 Colleen V. Chien, Inequality, Innovation, and Intellectual Property __ (Santa Clara Univ. Working Paper __, 2018) (on file with the author).
262 Initial exploratory work suggests that this may be an issue. *See also* Bessen, *supra* note __ (finding that small entities maintain their patents less than do large entities—implying that their patents are less valuable or that there is a liquidity problem that prevents small entities from capturing the value of their patents, even when obtained.)

access not only innovation but innovator details through the patent record. Leveraging these developments, this Article has described and applied novel empirical methods for studying innovators that patent, to reveal broad shifts in their profile, settings, and locations of over the past four decades, and demonstrated, through three mini-case studies, how improving our understanding of innovators can improve innovation.

APPENDIX

Table A: The Differential Treatment of Classes of Innovators Over Time

Law	Summary/State of the Law	Legal Provision
The Patent Act of 1790	All could apply for patents	Anyone who invented or discovered "any useful art, manufacture, engine, machine, or device, or any improvement therein not before known or used;[263] "he, she, or they" could apply for patents.
The Patent Act of 1793	U.S. citizens ("free white persons") could apply for patents	"[C]itizen or citizens of the United States"[264] could apply for patents. Under the Immigration and Naturalization Act of 1790, citizenship was reserved exclusively to "free White Persons"[265]
Act of April 17, 1800	U.S. citizens and foreigners resident for two years could apply for patents	In addition to citizens, "[a]ll aliens who at the time of petitioning [] shall have resided for two years within the United States" [266] could apply for patents.
Act of 1832	U.S. citizens and foreign residents intending to become citizens could apply for patents.	Alien residents who signed an oath attesting to their intention to become citizens could apply for patents; those who did not work their patents within a year of grant had their patents revoked[267]

---

[263] Patent Act of 1790, ch. 7, § 1, 1 Stat. 109 [hereinafter 1790 Act].

[264] Patent Act of 1793, ch. 11, § 1, 1 Stat. 318 [hereinafter 1793 Act].

[265] Naturalization Act of 1790, ch. 3, § 1, 1 Stat. 103. This excluded naturalization of Asians, American Indians, and free black immigrants. According to Haney-Lopez, this racial prerequisite to citizenship remained in force until 1952. *See* IAN HANEY LOPEZ, WHITE BY LAW: THE LEGAL CONSTRUCTION OF RACE 1 (New York Univ. Press rev. ed. 2006).

[266] Act of April 17, 1800, ch. 25, § 1, 2 Stat. 37, 38.

[267] *See Continental Paper Bag Co. v. Eastern Paper Bag Co.*, 210 U.S. 405, 429 (1908) (speaking of the right to patent, that "[t]he only qualification ever made was against aliens, in the act of 1832. That act extended the privilege of the patent law to aliens, but required them 'to introduce into public use in the United States the invention or improvement within one year from the issuing thereof,' and indulged no intermission of the public use for any period longer than six months. A violation of the law rendered the patent void. The act was repealed in 1836.") The actual language of the statue was a codification of the Supreme Court's decision in *Grant v. Raymond*, 31 U.S. (6 Pet.) 218 (1832).

Table A (cont'd): The Differential Treatment of Classes of Innovators Over Time

Law	Summary/State of the Law	Legal Provision
Patent Act of 1836	U.S. and foreign citizens (but not African-Americans) could apply for patents, foreigners paid higher fees.	Citizens and alien citizens[268] could apply for patents. The Supreme Court's *Dred Scott* decision in 1857 excluded "persons of African descent," free or slave, from U.S. citizenship.[269] U.S. citizens and resident aliens that promised to become citizens within a year paid an application fee of $30; British nationals paid $500, and all other foreigners, $300.[270]
Act of July 8, 1870	U.S. citizens and those who were about to become U.S. citizens could apply for patents	Designers of "new and original fabrics," that were or were about to become U.S. citizens[271] could apply for design patents.
1930 Tariff Act	Exclusion orders against infringing imports are available for patentholders with "domestic industries"	Patentholding complainants with "domestic industries"[272] are entitled to apply for exclusion orders against infringing imports.[273]
1952 Patent Act	Foreign inventive activity, unless published down, not considered for the purposes of determining prior art whereas U.S.	Only foreign printed publications count as prior art; domestic knowledge, public use, sale or printed publication count.[274] This changed when the United States joined the WTO in 1995 and, like other members, was required to treat citizens of other member countries as well or better than its own citizens under the principle of national treatment.[275] In 2011, as part of the

[268] Patent Act of 1836, ch. 357, §§ 6 & 9, 5 Stat. 117, 119 & 12 [hereinafter 1836 Act] (specifying that each patent applicant was to provide an oath describing, among other things, "of what country he is a citizen," as well as contemplating applicants could be "a citizen of the United States, or an alien.")

[269] *Dred Scott v. Sandford*, 60 U.S. (19 How.) 393 (1857) (finding that persons of African descent cannot be, nor were ever intended to be, citizens under the U.S. Constitution). As Frye recounts, Dred Scott not only precluded free blacks from rights to their inventions, but also precluded their slave owners, who could not take an oath attesting to be the "inventors" of their slaves' inventions, from such rights as well, defying the claim that slave owners "owned" slaves and their ideas. Frye, *supra* note ___, at 1-2.

[270] 1836 Act § 6; *see also* § 12 (limiting the filing of a caveat, an instrument similar to a patent, to citizens and aliens intending to become citizens).

[271] Act of July 8, 1870, ch. 230, §§ 40, 71, 16 Stat. 198, 203-04, & 209-10, *repealed by* Act of May 9, 1902, ch. 783, 32 Stat. 193.

[272] *Id. See also* Colleen Chien, *Patently Protectionist? An Empirical Analysis of Patent Cases at the International Trade Commission*, 50 WM. & MARY L. REV. 63 (finding that despite the domestic industry requirement, based on an empirical analysis, that domestic defendants have found themselves targeted almost as often as foreign defendants, most often together with them, but that domestic plaintiffs still use the venue to a greater degree).

[273] Pursuant to Section 337 of the Tariff Act of 1930, 19 U.S.C. § 1337 (2012).

[274] 35 U.S.C. §§ 102(a)-(b) (1952).

[275] *See* TRIPS. For a summary of how joining TRIPS resulted in a change to U.S. novelty rules, see https://www.uspto.gov/web/offices/com/doc/uruguay/SUMMARY.html.

	knowledge and use considered prior art.	America Invents Act, equal treatment was extended to all countries,[276]

[276] *Cf.* 35 U.S.C. § 102(b) (2006) & 35 U.S.C. § 102(a)(2) (2012).

Table A1: Patent Dataset Description

Dataset	Sample (Months per year)	Data Source	Number of Patents Analyzed	Number of Inventors Analyzed
1870 Patents	All patents	Early Patent Lists[277]	10616	10616 (for locations)
USPTO Entity Status	1.5% random sample	USPTO OCE	50708	Analyzed owners on 43127 patents
1896-1966 Patent Locations	0.5% random sample	Coded by Mechanical Turk	~1000	~1000
1906-1966 Patents	March-May	Innography	50327	50327 (First named inventor)
1976 Worldwide Patents	April-May	PatentsView	13,472	
1986 U.S. Inventor	March, September	PatentsView dataset in Bigquery	12,934	11,446[278]
1996 U.S. Inventor Patents	March, September	PatentsView dataset in Bigquery	17,746	20,308
2006 U.S. Inventor Patents	March, September	PatentsView dataset in Bigquery	31,451	40,634
2016 U.S. Inventor Patents	March 1-22[279]	PatentView dataset in Bigquery	22,806	32,167
1986-2016 Totals			84,937	104,555

277 *Utility Patents, Historical, Regional, and Specialized Patent and Trademark Research*, PAT. & TRADEMARK RES. CTR. ASS'N, http://ptrca.org/history (last visited Nov. 5, 2017).

278 For 1986-2016 U.S. patents, numbers represent numbers of unique inventors analyzed.

279 To be updated when the BigQuery data is updated.

Table A2: Dataset Description and Coding Coverage

Small Entity Dataset (2000-2015 Patents)				
Labels	Blanks	Large	Small/Micro	Not profiled
Total	7581	35927	7200	15%
Gender Classifier Coverage Rates Among U.S. Inventors				
Year	Social Security Classifier	Genderize.io	Agreement Rate	
1986	93.73%	99.98%	98.70%	
1996	91.36%	99.96%	98.56%	
2006	86.40%	99.98%	98.35%	
2016	81.03%	99.99%	97.80%	
Total	86.51%	99.98%	98.27%	
Ethnicity Classifier Coverage Rates Among U.S. Inventors				
Year	NamePrism	Census Classifier	Agreement Rate	Combined Coverage Rate
1986	82.54%	82.35%	98.19%	97.00%
1996	81.63%	83.07%	98.02%	96.67%
2006	77.19%	82.06%	97.92%	96.21%
2016	78.15%	80.57%	97.61%	96.35%
Total	78.93%	81.83%	97.88%	96.43%

Table A4 : 2014 State College and Patent Rankings [280]

States with the most colleges and universities per capita	States with the most colleges and universities	States with the most patent grant per capita	States with the most grants
Montana	California	California	California
New Jersey	New York	North Carolina	Texas
Nevada	Texas	Connecticut	New York
Texas	Pennsylvania	Pennsylvania	Washington
Maryland	Florida	Virginia	Massachusetts
Michigan	Ohio	Texas	Illinois
Idaho	Illinois	Georgia	Michigan
Florida	North Carolina	New Jersey	Florida
Delaware	Georgia	Iowa	Colorado
California	Virginia	Maryland	Ohio

[280] Author's analysis using data from U.S. CENSUS BUREAU, https://www.census.gov/ (last visited Jan. 27, 2017) (for population data); PatentsView, *supra* note 72; NAT'L CTR. FOR EDUC. STATISTICS, U.S. DEP'T OF EDUCATION, DIGEST TABLE 317.20: DEGREE-GRANTING POSTSECONDARY INSTITUTIONS, BY CONTROL AND LEVEL OF INSTITUTION AND STATE OR JURISDICTION: 2014-15 (Oct. 2015), https://nces.ed.gov/programs/digest/d15/tables/dt15_317.20.asp (in descending order: California, New York, Texas, Pennsylvania, Florida, Ohio, Illinois, North Carolina, Georgia, Virginia, Missouri, Massachusetts, Minnesota, Michigan, Tennessee).

Table A5: States Ranked By Patent Grant Densities

Grant Densities (patents/10K residents)							Annual Patent Grants						
2016 Rank	State	1976	1986	1996	2006	2016 [281]	2016 Rank	State	1976	1986	1996	2006	2016
1	CA	22.7	16.2	27.6	47.2	77.1	1	CA	5965	5252	1051 3	22295	23052
2	NC	7.5	8.0	18.2	30.6	42.6	2	TX	1935	2359	4173	6289	4933
3	CT	22.0	20.1	27.4	29.6	34.5	3	NY	4400	3153	5305	5985	4662
4	PA	33.0	24.6	29.7	27.1	34.2	4	WA	515	586	1179	3324	3435
5	VA	13.3	9.4	14.8	17.8	31.0	5	MA	1861	1484	2477	3973	3169
6	TX	15.2	14.2	20.6	24.9	30.1	6	IL	3543	2375	3144	3266	2743
7	GA	3.4	3.8	8.2	11.3	22.3	7	MI	2338	2082	3175	3774	2631
8	NJ	31.2	18.8	18.2	16.3	17.7	8	FL	916	1183	2088	2606	2601
9	IA	4.6	3.3	3.7	4.5	12.3	9	CO	523	512	1178	2161	2320
10	MD	7.9	5.3	8.8	10.5	12.0	10	OH	2731	2120	2604	2611	2312
11	CO	1.7	1.6	3.6	6.1	11.8	11	MN	829	902	1679	2689	2106
12	WI	4.2	4.3	7.2	9.4	11.7	12	NJ	3705	2756	3111	3197	2016
13	WA	1.1	1.2	2.3	6.0	10.9	13	PA	3122	2407	2929	2885	1980
14	OH	9.7	6.5	7.9	7.3	10.8	14	NC	482	534	1167	1985	1766
15	SC	3.5	3.4	6.7	7.4	10.4	15	GA	303	404	975	1485	1751
16	MA	4.5	3.3	4.9	7.1	9.6	16	CT	1523	1282	1475	1687	1269
17	MI	5.9	5.0	6.8	7.3	8.8	17	AZ	423	612	1060	1727	1224
18	IL	6.6	4.4	5.4	5.2	7.6	18	WI	797	808	1304	1709	1183
19	AL	5.6	4.8	4.6	5.4	7.4	19	IA	391	327	442	663	1119
20	NY	4.1	2.9	4.7	5.2	7.3	20	IN	1075	851	1288	1172	1110

Codes for Replication

[281] Projected based on actuals through July 15, 2016.

- DataCleaning Protocols: https://github.com/Tech-Leaderboard/data-cleaners. Correction of certain obvious spelling mistakes, like $teven for Steven, was not necessary because the mistaken entries were still recognized by the classifiers we used. Other corrections, like the substitution of a middle name for a first name when only an initial was available for a first name, were also made by the classifiers.
- Queries for patent extract on BigQuery: [add from Johnny]

FIG. 3A: Estimated Female Shares of US Inventors[282]

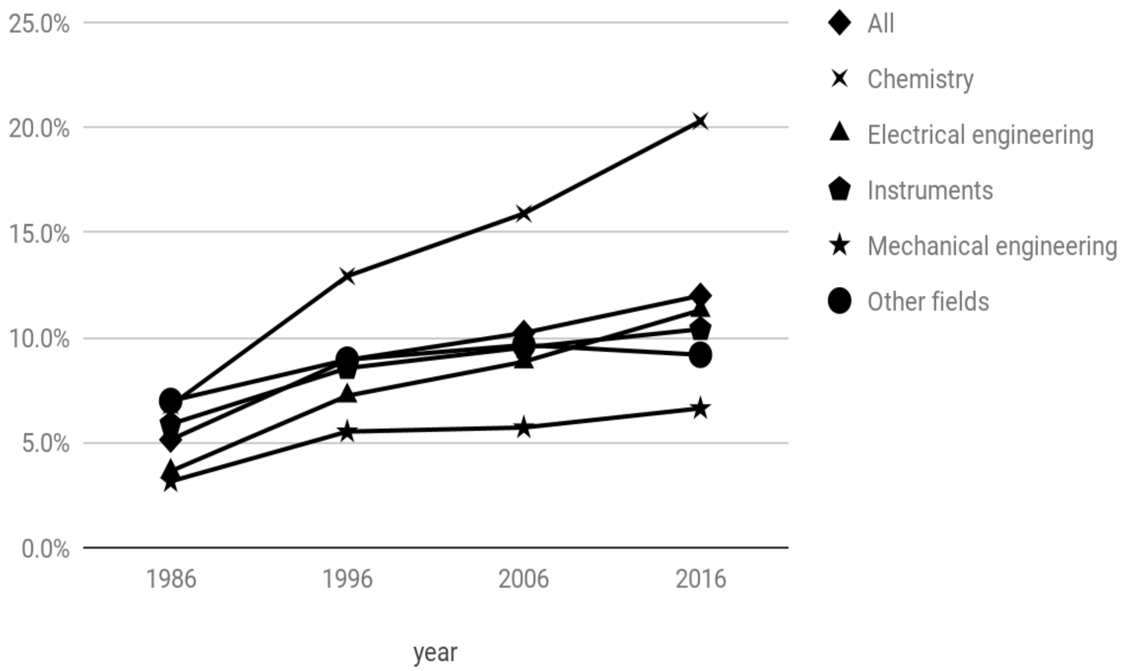
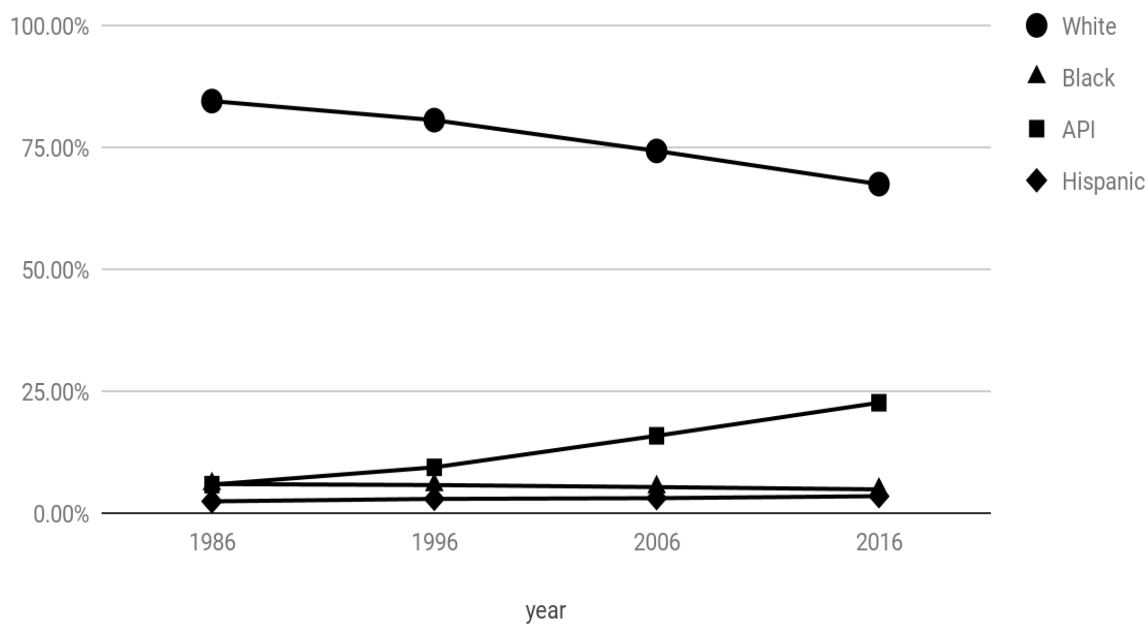


Fig. 3B: Estimated Ethnic Shares of US Inventors[283]

---

[282] For point values, see Appendix Table 3A.
[283] Values displayed based on cumulative probability.


100.00%
75.00%
50.00%
25.00%
0.00%
1986
1996
2006
2016
year
White
Black
API
Hispanic