Who Becomes an Inventor in America? The Importance of Exposure to Innovation*

Alex Bell, Harvard University
Raj Chetty, Harvard University and NBER
Xavier Jaravel, London School of Economics
Neviana Petkova, Office of Tax Analysis, US Treasury
John Van Reenen, MIT and Centre for Economic Performance

August 2018

Abstract

We characterize the factors that determine who becomes an inventor in the United States, focusing on the role of inventive ability ("nature") vs. environment ("nurture"). Using de-identified data on 1.2 million inventors from patent records linked to tax records, we first show that children's chances of becoming inventors vary sharply with characteristics at birth, such as their race, gender, and parents' socioeconomic class. For example, children from high-income (top 1%) families are ten times as likely to become inventors as those from below-median income families. These gaps persist even among children with similar math test scores in early childhood – which are highly predictive of innovation rates – suggesting that the gaps may be driven by differences in environment rather than abilities to innovate. We then directly establish the importance of environment by showing that exposure to innovation during childhood has significant causal effects on children's propensities to invent. Children whose families move to a high-innovation area when they are young are more likely to become inventors. These exposure effects are technology-class and gender specific. Children who grow up in a neighborhood or family with a high innovation rate in a specific technology class are more likely to patent in *exactly the same* class. Girls are more likely to invent in a particular class if they grow up in an area with more women (but not men) who invent in that class. These gender- and technology class-specific exposure effects are more likely to be driven by narrow mechanisms such as role model or network effects than factors that affect general human capital accumulation, such as the quality of schools. Consistent with the importance of exposure effects in career selection, women and disadvantaged youth are as under-represented among high-impact inventors as they are among inventors as a whole. These findings suggest that there are many "lost Einsteins" – individuals who would have had highly impactful inventions had they been exposed to innovation in childhood – especially among women, minorities, and children from low-income families.

---

*A preliminary draft of this paper was previously circulated under the title "The Lifecycle of Inventors." The opinions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Internal Revenue Service, U.S. Department of the Treasury, or the National Institutes of Health. We would particularly like to thank Philippe Aghion, with whom we started thinking about these issues, for inspiration and many insightful comments. We would like to also thank Daron Acemoglu, Ufuk Akcigit, Olivier Blanchard, Erik Hurst, Danny Kahnemann, Pete Klenow, Henrik Kleven, Richard Layard, Eddie Lazear, Josh Lerner, Alex Olssen, Jim Poterba, Scott Stern, Otto Toivanen, Heidi Williams, and numerous seminar participants for helpful comments and discussions. Trevor Bakker, Augustin Bergeron, Mike Droste, Jamie Fogel, Nikolaus Hidenbrand, Alexandre Jenni, Benjamin Scuderi, and other members of the Equality of Opportunity Project research team provided outstanding research assistance. This research was funded by the National Science Foundation, the National Institute on Aging Grant T32AG000186, the Lab for Economic Applications and Policy at Harvard, the European Research Council, the Economic and Social Research Council at CEP, the Washington Center for Equitable Growth, the Kauffman Foundation, the Bill and Melinda Gates Foundation, and the Robert Wood Johnson Foundation.

I Introduction

Innovation is widely viewed as a central driver of economic growth (e.g., Romer 1990, Aghion and Howitt 1992). As a result, many countries use a wide variety of policies to spur innovation, ranging from tax incentives to investments in education. Most existing work analyzing the effectiveness of such policies has examined their impacts on the rate of innovation at the firm, industry, or macroeconomic level (e.g., Becker 2015). In this paper, we take a different approach, focusing on the *individuals* who become inventors. By analyzing the factors that determine who becomes an inventor, we identify new approaches to increasing rates of innovation, especially among subgroups that are currently under-represented in the innovation sector.[1]

Although there is a growing body of work studying the backgrounds of inventors using historical data from the U.S. and contemporary data from Scandinavian countries (e.g., Khan and Sokoloff 1993, Akcigit et al. 2017, Aghion et al. 2017), relatively little is known about the individuals who become inventors in the modern era in the U.S. This is because most sources of data on innovation (e.g., patent records) do not record even basic demographic information, such as an inventor's age or gender.

We present a comprehensive portrait of inventors in the United States today by linking patent records to income tax records. Following standard practice in prior work on innovation, we define an "inventor" as an individual who holds a patent.[2] We link data on the universe of patent applications and grants in the U.S. between 1996 and 2014 to federal income tax returns to construct a panel dataset covering 1.2 million inventors (patent applicants or recipients). Using this new dataset, we track inventors' lives from birth to adulthood to identify factors that determine who becomes an inventor, focusing on the role of inventive ability ("nature") vs. environment ("nurture").[3] We organize our analysis into three parts.

In the first part of the paper, we show that children's characteristics at birth – their socioe-

[1]For example, it is important to understand whether the "extensive margin" decision to become an inventor is driven primarily by financial incentives or by non-financial factors such as the environmental "exposure effects" we investigate below. More broadly, studying who becomes an inventor also sheds light on the link between inequality and innovation and the mechanisms that drive career choice.

[2]The use of patents as a proxy for innovation has well-known limitations (e.g. Griliches 1990, OECD 2009). In particular, not all innovations are patented and not all patents are meaningful innovations. We address these measurement issues by showing that (a) our results hold if we focus on highly-cited (i.e., high-impact) patents and (b) the mechanisms that lead to the differences in rates of patenting across subgroups that we document are unlikely to be affected by these concerns.

[3]There is no sharp dichotomy between nature and nurture because behavior is likely determined by an interaction between the two factors, as emphasized e.g. in the literature on epigenetics. We therefore focus not on decomposing the relative importance of these two factors but on investigating whether and how environmental factors influence rates of innovation.

conomic class, race, and gender – are highly predictive of their propensity to become inventors. Children born to parents in the top 1% of the income distribution are ten times as likely to become inventors as those born to families with below-median income.[4] Whites are more than three times as likely to become inventors as blacks. And 82% of 40-year-old inventors today are men. This gender gap in innovation is shrinking gradually over time, but at the current rate of convergence, it will take another 118 years to reach gender parity.

Why do rates of innovation vary so sharply based on characteristics at birth? One potential explanation is that the differences stem from inherited differences in talents or preferences to pursue innovation as a career. An alternative explanation is that children from different backgrounds grow up in different environments and therefore end up pursuing different careers.

As a first step toward evaluating whether differences in inherited abilities can explain gaps in innovation, we use math test scores in early childhood as an (imperfect) proxy for innovative potential. We obtain data on test scores from 3rd to 8th grade by linking school district records for 2.5 million children who attended New York City public schools to the patent and tax records. Math test scores in 3rd grade are highly predictive of patent rates, but account for less than one-third of the gap in innovation between children from high- vs. low-income families.[5] This is because children from lower income families are much less likely to become inventors even conditional on having test scores at the top of their 3rd grade class. Differences in 3rd grade math scores also explain a small share of the gap in innovation by race, and virtually none of the gap in innovation by gender.

The gap in innovation explained by test scores grows in later grades, consistent with prior evidence that test score gaps widen as children progress through school (e.g., Fryer and Levitt 2004, Fryer 2011). Half of the gap in innovation by parent income can be predicted by differences in math test scores in 8th grade. Furthermore, gaps in innovation by parental income are relatively small conditional on the college that a child attends. These results suggest that low-income children start out on relatively even footing with their higher-income peers in terms of innovation ability, but fall behind over time, perhaps because of differences in their childhood environment. However,

[4]This pattern is not unique to innovation: children from high-income families are also substantially more likely to enter other high-skilled professional occupations and, more generally, reach the upper-tail of the income distribution. We focus on innovation here because it is thought to have particularly large positive social spillovers and because focusing on innovation has methodological advantages in understanding the mechanisms underlying career choice, as we discuss below.

[5]Although test scores in English are highly predictive of propensities to invent unconditionally, they have no predictive power conditional on test scores in math. This suggests that tests in early childhood are diagnostic of the specific skills that matter for innovation.

they do not provide conclusive evidence about the role of environment because test scores are an imperfect measure of inventive ability. If a child's ability to innovate is poorly captured by standardized tests, particularly at early ages, ability could still account for a substantial share of gaps in innovation.[6] Moreover, this analysis leaves open the possibility that differences in inherited preferences explain gaps in innovation.

In the second part of the paper, we study the impacts of childhood environment directly to address these issues. We show that exposure to innovation during childhood through one's family or neighborhood has a significant *causal* effect on a child's propensity to become an inventor.[7] We establish this result – which we view as the central empirical result of the paper – in a series of steps. We first show that children who grow up in commuting zones (CZs) with higher patent rates are significantly more likely to become inventors, even conditional on the CZ in which they work in adulthood. We then show this pattern holds not just for whether a child innovates but also for the technology category in which he or she innovates. For example, among people living in Boston, those who grew up in Silicon Valley are especially likely to patent in computers, while those who grew up in Minneapolis – which has many medical device manufacturers – are especially likely to patent in medical devices. We find similar patterns at the family level: children whose parents or parents' colleagues hold patents in a technology class are more likely to patent in exactly that field themselves.

These patterns of transmission hold even across the 445 narrowly defined technology subclasses into which patents can be classified. For example, a child whose parents hold a patent in amplifiers is much more likely to patent in amplifiers himself than in antennas. Moreover, the patterns are gender-specific: women are much more likely to patent in a specific technology class if female workers in their childhood CZ were especially likely to patent in that class. Conditional on women's patent rates, men's patent rates have no predictive power for women's innovation. Conversely, men's innovation rates are influenced by male rather than female inventors in their area.

Under the assumption that differences in genetic abilities do not generate differences in propensities to innovate across narrow technology classes in a gender-specific manner, this set of results on patenting by technology class implies that exposure to innovation during childhood has a causal

[6] On the other hand, since children from different socioeconomic backgrounds are exposed to different environments even before they enter school, these calculations could overstate the portion of the gap in innovation that is due to differences in inventive ability.

[7] We use the term "exposure to innovation" to mean having contact with someone in the innovation sector, e.g. through one's family or neighbors. We do not distinguish between the mechanisms through which such exposure matters, which could range from specific human capital accumulation to changes in aspirations.

effect on the type of innovation one pursues. Intuitively, as long as genetics do not govern one's ability to invent an amplifier rather than an antenna in a gender-specific manner, the close alignment between the subfield in which children innovate and the type of innovation they were exposed to in their families or neighborhoods must be driven by causal exposure effects. Formally, the sharp variation in rates of innovation across technology classes and gender subgroups provides a set of overidentifying restrictions that allow us to distinguish exposure effects from plausible models of selection in observational data.

The technology class-level results discussed above show that exposure affects the type of innovation one pursues, but do not necessarily imply exposure affects whether one chooses to become an inventor to begin with. To test whether exposure affects the *level* of innovation, we study the outcomes of children whose families move across CZs, exploiting variation in the timing of moves between areas as in Chetty and Hendren (2018). We find that children who move to areas with higher rates of innovation (among adults) earlier in their childhood are more likely to become inventors themselves. Under the identifying assumption that unobservable determinants of children's outcomes in adulthood are uncorrelated with the age at which they move to a different area – an assumption validated by Chetty and Hendren (2018) – this result implies that neighborhoods have causal effects on the total level of innovation. The estimates imply that approximately 75% of the observational correlation between children's propensity to become inventors and patent rates among adults in their CZ is driven by causal effects of environment. It follows that moving a child from a CZ that is at the 25th percentile of the distribution in terms of the fraction of adult inventors (e.g., New Orleans, LA) to the 75th percentile (e.g., Austin, TX) would increase his or her probability of becoming an inventor by 37%.

The exposure effects we document here are consistent with recent evidence documenting neighborhood exposure effects on earnings, college attendance, and other outcomes (Chetty et al. 2016). Such neighborhood effects have typically been attributed to factors that affect general human capital accumulation, such as the quality of local schools or residential segregation. Our findings show that, at least in the context of innovation, such mechanisms are unlikely to be the sole reason that childhood environment matters, as it is implausible that some neighborhoods prepare children to innovate in one particular technology class such as amplifiers. Rather, they point to mechanisms such as transmission of specific human capital, mentoring, or networks (e.g., through internships) that lead children to pursue certain career paths. Children from low-income families, minorities, and women are less likely to have such exposure through their families and neighborhoods, which helps explain why they have significantly lower rates of innovation overall. For example, our estimates imply that if girls were as exposed to female inventors as boys are to male inventors in their childhood CZs, the current gender gap in innovation would shrink by half.

In the final section of the paper, we briefly examine inventors' career trajectories, focusing on how the returns to innovation vary across subgroups to learn about which types of individuals appear to be screened out of innovation.[8] We find that inventors from under-represented groups (women, minorities, and those from low-income families) have very similar earnings and citations to other inventors on average. Put differently, women and disadvantaged youth are just as under-represented among star inventors as they are among inventors as a whole. This result is consistent with our finding that exposure is a central determinant of innovation. A lack of exposure may prevent some individuals ("lost Einsteins") from pursuing a career in innovation even though they would have had highly impactful innovations had they done so. Hence, drawing more children from under-represented groups into careers in innovation can have substantial impacts not only on the total number of inventors but also on the number of high-impact, high-return inventions.

We conclude that increasing exposure to innovation among children who (a) excel in math and science at early ages and (b) are from under-represented groups can have large impacts on aggregate innovation. Indeed, we estimate that if women, minorities, and children from lower-income families were to invent at the same rate as white men from high-income (top-quintile) families, the total number of inventors in the economy would quadruple.[9]

Our results do not, however, provide guidance on which specific policies are most effective in increasing exposure to innovation. To facilitate future work evaluating such policies, we construct a set of publicly available data tables that provide statistics on patent rates and citations by technology category, parent income group, gender, age, commuting zone, and college. In addition, we report statistics on inventors' income distributions by year and citations. These statistics can be used to study a variety of issues, ranging from the impacts of local economic conditions and policies on rates and types of innovation to how the returns to innovation have changed over time.

Related Literature. Our results build on and contribute to several literatures. First, our results relate to the literature on career choice (e.g., Topel and Ward 1992, Hall 2002). Some studies in this literature have used data on specific occupations – such as medicine and law – to show that

[8]We present a more comprehensive analysis of inventors' labor market careers in the working paper version of this study (Bell et al. 2017).

[9]We caution that our results do not necessarily imply that aggregate welfare would be higher if these "lost Einsteins" were to enter innovation, as these individuals might currently be pursuing other careers that also have substantial social returns.

children are particularly likely to pursue their parents' occupations (e.g., Laband and Lentz 1983, Lentz and Laband 1989), but they have not separated causal exposure effects from selection effects as we do here. While the mechanisms we document may apply to other careers as well, we focus on innovation because of its importance for economic growth (e.g., Jones and Williams 1999, Bloom et al. 2013).

Second, our results relate to the literature on the misallocation of talent across occupations (e.g., Murphy et al. 1991, Hsieh et al. 2016). Our analysis does not directly show that talent is misallocated, but our finding that the allocation of talent to innovation is driven partly by differences in exposure rather than inherited abilities is consistent with the premise of this literature. Indeed, our results raise the possibility that the welfare costs of distortions in the allocation of talent may be even greater than predicted by models such as Hsieh et al. (2016), since some of the individuals who fail to pursue innovation due to a lack of exposure are superstars rather than marginal entrants. More broadly, our findings suggest that improving opportunities for children from low-income or minority backgrounds (e.g., Heckman 2006, Card and Giulano 2014) could increase not just their own earnings but also economic growth by improving the allocation of talent.

Third, our study contributes to the nascent literature on the origins of inventors discussed above that sheds light on the "supply" of innovation (Goolsbee 1998, Romer 2000). For example, Aghion et al.'s (2017) study of inventors in Finland documents gaps in innovation by parental background consistent with our results and characterizes the predictive power of other factors that we do not observe in our data, such as IQ and parental education.[10] Our study also contributes to a related literature on the determinants of entrepreneurship that analyzes the role of ability (Nicolaou et al. 2008, Shane and Nicolaou 2013) and peer effects (Giannetti and Simonov 2009, Nanda and Sørensen 2010). Our analysis complements these studies by (a) identifying different factors that affect career choice, most importantly the causal effect of childhood exposure and (b) presenting comprehensive data and publicly available statistics on inventors' origins and careers in the United States.

The paper is organized as follows. Section II describes the data. Section III presents the results on inventors' characteristics at birth. Section IV analyzes the role of childhood environments. Section V presents results on inventors' career trajectories. Section VI concludes. Data tables on patent rates by subgroup can be downloaded from the Equality of Opportunity Project website.

[10] Other recent studies in a similar vein include Giuri et al. (2007), Nicholas (2010), Azoulay et al. (2011), Toivanen and Vaananen (2012), Dorner et al. (2014), Jung and Ejermo (2014), and Lindquist et al. (2015). A forerunner of this recent work was a classic study by Schmookler (1957) of 57 American inventors.

II Data

In this section, we describe our data sources, define the samples and key variables we use in our analysis, and present summary statistics.

II.A Data Sources

Patent Records. We obtain information on patents from two sources. First, we use information on patent grants from a database hosted by Google, which contains the full text of all patents granted in the U.S. from 1976 to present. We focus on the 1.7 million patents that were granted between 1996 and 2014 to U.S. residents. Second, we use data on 1.6 million patent applications between 2001 and 2012 provided by Strumsky (2014).[11]

We define an individual as an inventor if he or she is listed as an inventor on a patent application between 2001-2012 or grant between 1996-2014; for simplicity, we refer to this outcome as "inventing by 2014" below. Importantly, we include all individuals listed as inventors, not just those assigned intellectual property rights. In particular, inventors employed by companies are listed as inventors, while their company is typically listed as the assignee. In addition to inventors' names, we also extract information on inventors' geographic location (city and state) when they filed the patent and the 3-digit technology class to which the patent belongs, as assigned by the United States Patent and Trademark Office (USPTO). We classify patents into technology categories using the classification developed in the NBER Patent Data Project by Hall et al. (2001). We assign each inventor in our data a single technology class based on the class in which he or she has the most patents, breaking ties randomly. We obtain data on the number of times each granted patent was cited from its issuance date until 2014 from the USPTO's full-text issuance files.

Tax Records. We use federal income tax records spanning 1996-2012 to obtain information such as an individual's gender and age, geographic location, and own and parental income. The tax records cover all individuals who appear in the Death Master file produced by the Social Security Administration, which includes all persons in the U.S. with a Social Security Number or Individual Taxpayer Identification Number (ITIN). The data include both income tax returns (1040 forms) and third-party information returns (e.g., W-2 forms), which give us information on the earnings of those who do not file tax returns.

[11] In 2001, the U.S. began publishing patent *applications* (and not just patent grants) 18 months after filing. For a fee, applicants can choose to have their filing kept secret; 15% of applicants choose to do so. To ensure that this missing data problem does not generate selection bias, we verify that the results we report below are all robust to defining inventors purely using patent grants rather than applications.

The patent data were linked to the tax data using an inventor's name, city, and state. In the tax data, these fields were obtained from the Death Master file, 1040 forms, and third-party information returns (see the Online Appendix for a complete description of the matching procedure). 88% of individuals who applied for or were granted a patent were successfully linked, with higher match rates in more recent years since information returns are unavailable prior to 1999.

We evaluate the quality of our matching algorithm by using external data on ages for a subset of inventors from Jones (2010). The age of the inventor recorded in the Death Master file matches the age reported in Jones's dataset in virtually all cases, confirming that our algorithm generates virtually no false matches. The 12% of inventors who are not matched are individuals with common names that are difficult to link to unique records (e.g., "John Smith"), individuals with spelling errors in their names or addresses, or individuals who listed different addresses on their patent applications and tax forms. The observable characteristics (in the patent data) of unmatched inventors are very similar to those of those of matched inventors, suggesting that the individuals we match are representative of inventors in the U.S.

New York City School District Records. We use data from the New York City (NYC) school district to obtain information on test scores in childhood for the subset of individuals who attended New York City public schools. These data span the school years 1988-1989 through 2008-2009 and cover roughly 2.5 million children in grades 3-8. Test scores are available for English language arts and math for students in grades 3-8 in every year from the spring of 1989 to 2009, with the exception of 7th grade English scores in 2002. These data were linked to the tax data by Chetty et al. (2014a) with an 89% match rate, and we use their linked data directly in our analysis.

After these three databases were linked, the data were de-identified (i.e., individual identifiers were removed) and the analysis was conducted using the de-identified dataset.

II.B Sample Definitions

We use three different samples in our empirical analysis: full inventors, intergenerational, and New York City schools.

Full Inventors Sample. Our first analysis sample consists of all inventors (individuals with patent grants or applications) who were successfully linked to the tax data. There are approximately 1.2 million individuals in this sample. This sample is structured as a panel from 1996 to 2012, with data in each year on individual's incomes, patents, and other variables. We use this sample to analyze inventors' labor market careers in Section V.

Intergenerational Sample. Much of our empirical analysis compares inventors to non-inventors in terms of characteristics at birth (Section III) and childhood environment (Section IV). To measure conditions at birth and childhood location, we must link individuals to their parents. To do so, we use the sample constructed by Chetty et al. (2014b) to study intergenerational mobility, focusing on all children in the tax data who (1) were born in the 1980-84 birth cohorts, (2) can be linked to parents, and (3) were U.S. citizens as of 2013. Chetty et al. (2014b, Appendix A) describe how this intergenerational sample is constructed starting from the raw tax data; here, we briefly summarize its key features.

We define a child's parents as the first tax filers between 1996 and 2012 to claim the child as a dependent and were between the ages of 15 and 40 when the child was born. Since children begin to leave the household after age 16, the earliest birth cohort that we can reliably link to parents is the 1980 birth cohort (who are 16 in 1996, when our data begin). Children are assigned parent(s) based on the first tax return on which they are claimed, regardless of subsequent changes in the parents' marital status or dependent claiming. Although parents who never file a tax return cannot be linked to children, we still identify parents for more than 90% of children, as the vast majority of children are claimed at some point because of the tax benefits of claiming children. We restrict the sample to children who are citizens in 2013 to exclude individuals who are likely to have immigrated to the U.S. as adults, for whom we cannot measure parent income. We cannot directly restrict the sample to individuals born in the U.S. because the database only records current citizenship status.[12]

Since few individuals patent in or before their early twenties, we focus on individuals in the 1980-84 birth cohorts, who are between the ages of 28-32 in 2012, the last year of our data. There are 16.4 million individuals in our primary intergenerational analysis sample, of whom 34,973 are inventors. To assess whether our results are biased by focusing on innovation at relatively early ages (by age 32), we also examine a set of older cohorts using data from Statistics of Income (SOI) cross-sections, which provide 0.1% stratified random samples of tax returns prior to 1996. The SOI cross-sections provide identifiers for dependents claimed on tax forms starting in 1987, allowing us to link parents to children back to the 1971 birth cohort (Chetty et al. 2014b, Appendix A). There are approximately 11,000 individuals, of whom 131 are inventors, in the 1971-72 birth cohorts in the SOI sample that we use to study innovation rates up to age 40.

[12]In addition, we limit the sample to parents with positive income (excluding 1.5% of children) because parents who file a tax return – as is required to link them to a child – yet have zero income are unlikely to be representative of individuals with zero income while those with negative income typically have large capital losses, which are a proxy for having significant wealth.

New York City Schools Sample. When analyzing whether test scores explain differences in rates of innovation (Section III), we focus on the sample of children in the NYC public schools data linked to the tax data. We also use this sample when analyzing differences in innovation rates by race and ethnicity, as race and ethnicity are only observed in the school district data. We focus on children in the 1979-1985 birth cohorts for the test score analysis because the earliest birth cohort observed in the NYC data is 1979. As in Chetty et al. (2014a), we exclude students who are in classrooms where more than 25% of students are receiving special education services and students receiving instruction at home or in a hospital. There are approximately 430,000 children in our NYC schools analysis sample, of whom 452 are inventors.

II.C Variable Definitions and Summary Statistics

In this subsection, we define the key variables we use in our analysis and present summary statistics. We measure all monetary variables in 2012 dollars, adjusting for inflation using the consumer price index (CPI-U).

Income. We use two concepts to measure individuals' incomes: wage earnings and total income. Wage earnings are total earnings reported on an individual's W-2 forms. Total (individual) income is wage earnings as well as self-employment income and capital income. Total income is defined for tax filers as Adjusted Gross Income (as reported on the 1040 tax return) plus tax-exempt interest income and the non-taxable portion of Social Security and Disability benefits minus the spouse's W-2 wage earnings (for married filers). For non-filers, total income is defined as wage earnings. Individuals who do not file a tax return and who have no W-2 forms are assigned an income of zero.[13] Because the database does not record W-2's and other information returns prior to 1999, we cannot reliably measure individual earnings prior to that year, and therefore measure individuals' incomes only starting in 1999. Income is measured prior to the deduction of individual income taxes and employee-level payroll taxes.

Parents' Incomes. Following Chetty et al. (2014b), we measure parent income as total pre-tax income at the household level. In years where a parent files a tax return, we define family income as Adjusted Gross Income (as reported on the 1040 tax return) plus tax-exempt interest income and the non-taxable portion of Social Security and Disability benefits. In years where a parent does not file a tax return, we define family income as the sum of wage earnings (reported on form W-2), unemployment benefits (reported on form 1099-G), and gross social security and disability benefits

[13] Importantly, these observations are true zeros rather than missing data. Because the database covers all tax records, we know that these individuals have no taxable income.

(reported on form SSA-1099) for both parents.[14] In years where parents have no tax return and no information returns, family income is coded as zero. As in Chetty et al. (2014b), we average parents' family income over the five years from 1996 to 2000 to obtain a proxy for parent lifetime income that is less affected by transitory fluctuations. We use the earliest years in our sample to best reflect the economic resources of parents while the children in our sample are growing up.

Geographic Location. In each year, individuals are assigned ZIP codes of residence based on the ZIP code from which they filed their tax return. If an individual does not file in a given year, we search W-2 forms for a payee ZIP code in that year. Non-filers with no information returns are assigned missing ZIP codes. We map ZIP codes to counties and CZs using the crosswalks and methods described in Chetty et al. (2014b, Appendix A). For children whose parents were married when they were first claimed as dependents, we always track the mother's location if marital status changes.

College Attendance. Chetty et al. (2017) construct a roster of attendance at all colleges in the U.S. from 1999-2013 by combining information from IRS Form 1098-T, an information return filed by colleges on behalf of each of their students to report tuition payments, with Pell Grant records from the Department of Education.[15] We assign each child in the intergenerational sample to the college he or she attends (if any) for the most years between ages 19-22. See Chetty et al. (2017, Appendix B) for further details on how colleges are identified.

Test Scores. We obtain data on standardized test scores directly from the New York City school district database. The tests were administered at the New York City school district level during the period we study. Following Chetty et al. (2014a), we normalize the official scale scores from each exam (math and English) to have mean zero and standard deviation one by year and grade to account for changes in the tests across school years.

Summary Statistics. Table I presents descriptive statistics for the three analysis samples described above. Column 1 presents statistics for the full inventors sample; columns 2 and 3 consider inventors and non-inventors in the intergenerational sample; and columns 4 and 5 consider inventors and non-inventors in the NYC schools sample.

In the full inventors sample, the median number of patent applications between 1996-2012 is 1

[14] Since we do not have W-2's prior to 1999, parent income is coded as 0 prior to 1999 for non-filers. Assigning non-filing parents 0 income has little impact on our estimates because only 3.1% of parents in the full analysis sample do not file in each year prior to 1999 and most non-filers have very low W-2 income (Chetty et al. 2014b). For instance, in 2000, the median W-2 income among non-filers in our baseline analysis sample was $0.

[15] All institutions qualifying for federal financial aid under Title IV of the Higher Education Act of 1965 must file a 1098-T form in each calendar year for any student that pays tuition. The Pell Grant records are used to identify students who pay no tuition.

and the median number of citations per inventor is also only 1. But these distributions are very skewed: the standard deviations of the number of patent applications and citations are 11.1 and 118.1, respectively. Inventors have median annual wage earnings of $83,000 and total income of $100,000. Again, these distributions are very skewed, with large standard deviations and mean incomes well above the medians. The mean age of inventors is 44 and 13% of inventors in the sample are women.

The intergenerational and NYC school samples have younger individuals because they are restricted to more recent birth cohorts. As a result, inventors in these subsamples have lower median incomes, patent applications, and citations than in the full sample.

III Inventors' Characteristics at Birth

In this section, we study how rates of innovation differ along three key dimensions determined at birth: parental income, race, and gender. We first document gaps in rates of innovation and then use test score data to assess the extent to which these gaps can be explained by differences in abilities to innovate.

III.A Gaps in Innovation by Characteristics at Birth

Parental Income. Figure Ia plots the fraction of children who invent by 2014 vs. their parents' income percentile using our intergenerational analysis sample (children in the 1980-84 birth cohorts). We assign parents percentile ranks by ranking them based on their mean household income from 1996 to 2000 relative to other parents with children in the same birth cohort. Children from higher-income families are significantly more likely to become inventors. 8 out of 1,000 children born to parents in the top 1% of the income distribution become inventors, 10 times higher than the rate among those with below-median-income parents. The relationship is steeply upward sloping even among high-income families: rates of innovation rise by 22% between the 95th percentile ($193,322) and 99th percentile ($420,028) of the parental income distribution. This pattern suggests that liquidity constraints or differences in resources are unlikely to fully explain why parent income matters, as liquidity constraints are less likely to bind at higher income levels and resources presumably have diminishing marginal returns.

Figure Ib shows that the probability a child has highly-cited patents – defined as having total citations in the top 5% of his or her cohort's distribution – has a very similar relationship to parental income. Hence, the relationship between patenting and parent income is not simply driven by children from high-income families filing low-value or defensive patents at higher rates. The pattern in Figure I also remains robust at older ages, allaying the concern that children from higher-income families may simply patent earlier than those from low-income families. In particular, using the Statistics of Income 0.1% sample, we find that the relationship between rates of innovation between ages 30 and 40 and parental income remains qualitatively similar (Online Appendix Figure Ia). Defining inventors purely on the basis of patent grants or patent applications also yields similar results (Online Appendix Figure Ib).

The relationship between innovation and parental income is representative of the relationship between achieving professional success and parental income more generally. Children's propensities to reach the upper tail of the income distribution have a similarly convex and sharply increasing relationship with parental income (Online Appendix Figure II). For instance, children with parents in the top 1% of the parent income distribution are 27 times more likely to reach the top 1% of their birth cohort's income distribution and 10.6 times more likely to reach the top 5% of their cohort's income distribution than those born to parents below the median. As discussed in the introduction, we focus on innovation here (rather than professional success in general) because of innovation's relevance for economic growth, its unique risk profile, and its advantages in characterizing mechanisms more precisely. However, the results and mechanisms we establish here may apply to other careers beyond innovation.

Race and Ethnicity. Next, we turn to gaps in innovation by race and ethnicity. Since we do not observe race or ethnicity in the tax data, we use the New York City school district sample for this analysis. The first set of bars in Figure II shows the fraction of children who patent by 2014 among white non-Hispanic, Black non-Hispanic, Hispanic, and Asian children. 1.6 per 1,000 white children and 3.3 per 1,000 Asian children who attend NYC public schools between grades 3-8 become inventors. These rates are considerably higher than those of Black children (0.5) and Hispanics (0.2), consistent with evidence from Cook and Kongcharoen (2010).[16]

Since there are significant differences in parental income by race and ethnicity, the raw gaps across race and ethnicity partly reflect the income gradient shown in Figure I. To separate these two margins, we control for differences in income by non-parametrically reweighting the parental

[16]The innovation rates are lower than those in Figure Ia because NYC public schools have predominantly low-income students, with more than 75% of students from families with incomes below the national median. NYC public schools also have a much larger share of minorities than the U.S. population: 19.5% of the children in our NYC sample are white, 9.6% are Asian, 33.7% are Hispanic, and 36.0% are Black. Although we cannot be sure that the racial patterns within the NYC schools hold nationally, we do find that the relationship between parental income and innovation in the NYC sample is very similar to the national pattern in Figure Ia, suggesting that it provides representative evidence at least on the socioeconomic dimension (Online Appendix Figure Ic).

income distributions of Blacks, Hispanics, and Asians to match that of whites in the NYC sample, following the methodology of DiNardo et al. (1996). We divide the parental income distribution of children in the NYC sample into ventiles (20 bins) and compute mean patent rates across the 20 bins for each racial/ethnic group, weighting each bin by the fraction of white children whose parents fall in that income bin (i.e., integrating over the income distribution for whites).

The second set of bars in Figure II plot the resulting innovation rates, which can be interpreted as the innovation rates that would prevail for each group if it had the same income distribution as whites. Adjusting for income differences does not eliminate the racial and ethnic gaps, but changes their magnitudes. The Black-white gap falls by a factor of 2 (from 1.1/1000 to 0.6/1000). The white-Asian gap widens from 1.7/1000 to 2.6/1000 when we reweight by income, as Asian parents in NYC public schools have lower incomes on average than white parents. The Hispanic-white gap remains essentially unchanged.

Gender. Finally, we examine gaps in innovation by gender. Since gender is recorded in the tax data for all individuals in the population, we use the full inventors sample for this analysis. The advantage of doing so is that we can examine gender differences in rates of innovation not just for those born in the 1980s as in our intergenerational sample, but for older cohorts as well.

Figure III plots the fraction of female inventors – individuals who applied for or were granted a patent between 1996 and 2014 – by birth cohort.[17] Consistent with prior work (Thursby and Thursby 2005, Ding et al. 2006, Hunt 2009, Kahn and Ginther 2017), we find substantial gender differences in innovation for those in the prime of their careers today; for instance, 18% of inventors born in 1980 are female. What is less well known from prior work is the rate at which this gap is changing over time. Figure III shows that the fraction of female inventors was only 7% in the 1940 cohort and has risen monotonically and linearly over time. However, the rate of convergence is slow: a 0.27 percentage point (pp) increase in the fraction of female inventors per cohort on average, based on a linear regression. At this rate, it will take another 118 years to reach gender parity in innovation.

[17]Because we examine patenting in a fixed time window, we measure patent rates at different ages for different cohorts, ranging from ages 56-72 for the 1940 cohort to ages 16-32 for the 1980 cohort. This approach yields consistent estimates of the gender gap across cohorts if gender differences in patenting do not vary by age. While we cannot evaluate the validity of this assumption across all cohorts, examining patent rates at a fixed age (e.g., age 40) over the 17 cohorts we can analyze yields similar results (not reported).

III.B Do Differences in Abilities Explain the Gaps in Innovation?

Why do rates of innovation vary so widely across individuals with different characteristics at birth? One potential explanation is that the differences stem from inherited differences in abilities to innovate or preferences to pursue innovation as a career.

In this subsection, we take a step toward evaluating the role of differences in abilities to invent by using data on childhood test scores for children in our New York City schools sample. Although students who attend New York City public schools are a selected subgroup, differences in innovation rates by parental income (Online Appendix Figure Ic) and gender (Table I) are very similar in the NYC school district sample as in the full intergenerational sample. We consider whether math test scores – an imperfect proxy for inventive ability that nonetheless proves to be highly predictive of innovation rates – can account for the gap in innovation within the NYC sample by income, race, and gender in turn.

Parental Income. In Table II, we estimate the fraction of the gap in innovation by parental income that can be predicted by math test scores in 3rd grade (the first grade we observe in the NYC data). We define "high-income" children as children with parents in the top income quintile within the NYC sample, placing all others in the "lower-income" category; using other thresholds to divide the two groups yields similar results. We focus on math test scores because scores in English do not predict innovation rates conditional on math scores (Online Appendix Table I).[18]

The first row of Table II shows that 1.93 out of 1,000 children from top-quintile families born between 1979-85 invent by 2014, as compared with 0.52 out of 1,000 children from lower-income families. The raw gap in innovation across these income groups is thus 1.41 inventors per 1,000 children. In the second row, we reweight the test scores of the lower-income students to match those of children from high income families, following the methodology of DiNardo et al. (1996) as in our analysis of income and race above. We divide the 3rd grade math test score distribution of children in the NYC sample into ventiles (20 bins) and compute mean patent rates across the 20 bins for the lower-income group, weighting each bin by the fraction of high-income children with test scores in that bin. The second row of Table II shows that, according to this statistical decomposition, children from lower-income families would have a patent rate of 0.96 per 1,000 (rather than 0.52) if they had the same test scores as children from high-income families. The patent rate rises because children from high-income families have higher test scores in 3rd grade; for instance, children from

[18] The same is not true for success on other dimensions: for instance, both math and English scores are predictive of the probability that a child reaches the top 1% of the income distribution (Online Appendix Table I).

the top income quintile score 0.65 SD higher on average than children from lower quintiles (Online Appendix Figure IIIa). However, these differences in test scores account for less than one third of the raw gap in innovation, as the gap remains at 0.97 per 1,000 even after adjusting for differences in test scores, as shown in column 3 of Table II.

Figure IVa illustrates why test scores fail to fully predict the gap in innovation by plotting innovation rates vs. test scores for children with parents in the top quintile (circles) and those with lower-income parents (triangles). Each point in this figure shows the fraction of inventors within a ventile of the test score distribution. In high-income families, children who score highly on 3rd grade math tests are much more likely to become inventors than those with lower test scores. By contrast, in lower-income families, children with higher test scores do not have much higher innovation rates. As a result, among students with test scores in the top 5% of the distribution, those from high-income families are more than twice as likely to become inventors as those from lower-income families. This result suggests that becoming an inventor in America relies on two traits: having high inventive ability (as proxied for by math test scores early in childhood) *and* being born into a high-income family.[19]

To obtain further insight into the role of inventive ability, we repeat the preceding analysis using test scores in later grades. Figure V plots the fraction of the raw gap in innovation that is accounted for by math test scores in each grade from grades 3-8. As children get older, test scores account for more of the gap in innovation by parental income. By 8th grade, 48% of the gap can be predicted by differences in test scores, significantly higher than the 31% in 3rd grade. Based on a linear regression across the six grades in which we observe scores, we estimate that on average an additional 3.2 percentage points of the gap is accounted for by test scores each year ($p < 0.01$).

Extrapolating linearly back to birth, our estimates imply that only 5.7% of the gap in innovation would be predicted by math test scores (our proxy for inventive ability) at birth. Conversely, test scores at the end of high school would account for 60.1% of the gap.[20] These results suggest that low-income children start out on even footing with their higher-income peers in terms of inventive

[19] This figure also implies that efforts to increase innovation among under-represented groups are likely to have the biggest impacts if they are targeted at children who excel in math and science at early ages. Since such efforts are unlikely to raise the innovation rates of children from under-represented groups beyond those observed for children with comparable test scores from advantaged backgrounds, Figure IVa suggests that there is limited scope to increase innovation rates among low-income children who score below the 90th percentile on math tests in 3rd grade. However, there may be substantial potential to do so among those who score in the top 10%.

[20] Naturally, the evolution of gaps in inventive ability may differ at earlier and later ages, so the results of these extrapolations should be interpreted with caution. We use these calculations simply to illustrate that the gaps in test scores expand sufficiently rapidly during childhood that they would account for essentially none of the gap in innovation if (hypothetically) measured at birth, but the majority of the gap if measured at the end of high school.

ability, but fall behind steadily as they grow older, perhaps because of differences in childhood environment.

Consistent with this conclusion, we find that gaps in innovation by parental income are relatively small among children who attend the same college. Figure VIa lists the ten colleges (among colleges with at least 500 students per cohort) whose students are most likely to become inventors.[21] Figure VIb presents a binned scatter plot of innovation rates vs. parental income rank among students at these ten high-innovation colleges. 7.1% of children with parents in the top 1% of the national income distribution become inventors at these colleges, compared with 4.0% of children from below-median-income families. This gap is an order of magnitude smaller than the 10 to 1 gap shown in Figure Ia for the nation as a whole, suggesting that children's levels of achievement around age 20 almost fully account for gaps in innovation. More broadly, this finding suggests that most of the innovation gap is explained by factors that affect children *before* they enter the labor market, as we show in Section IV.

Race and Ethnicity. We use analogous methods to those above to estimate how much of the racial gaps in innovation can be accounted for by test scores in the New York City schools sample. The third set of bars in Figure II show the innovation rates that would prevail if all children had 3rd grade math test scores comparable to those of whites. The gaps shrink modestly, showing that test scores account for very little of the racial gaps in innovation. For example, the Black-white gap shrinks from 1.1 to 1.0, a change of less than 10%, while the Asian-white gap falls by 9%. Figure IVb illustrates why this is the case by plotting patent rates vs. test scores by race and ethnicity. Even conditional on test scores, whites and Asians are substantially more likely to become inventors than Blacks and Hispanics. Very few of even the highest-scoring Black and Hispanic children pursue innovation.

Replicating the reweighting analysis by grade, we find that test scores in later grades account for more of the racial gaps in innovation, consistent with the patterns for income. For instance, 51% of the gap in patent rates between Asians and other racial and ethnic groups can be explained by 8th grade test scores.

Gender. Finally, we conduct an analogous exercise for gender, reweighting girls' test scores to match that of boys. Math test scores in 3rd grade account for only 2.4% of the difference

[21] Innovation rates for every college in the U.S. that has at least 10 inventors in our sample are provided in Online Data Table III. The college-level estimates are blurred to protect confidentiality using the procedure in Chetty et al. (2017, Appendix C). The degree of error due to the blurring procedure is smaller than the degree of sampling error in the estimates.

in innovation rates between males and females (Online Appendix Table II). This is because the distribution of math test scores for boys and girls is extremely similar in 3rd grade (Online Appendix Figure IIIb). Similar to the patterns by race and parental income, high-scoring girls are much less likely to become inventors than high-scoring boys (Figure IVc).

Even in 8th grade, test scores account for only 8.5% of the gender gap in innovation. One explanation for why the gender gap in test scores expands less across grades than racial and class gaps is that boys and girls attend similar schools and grow up in similar neighborhoods, whereas children with different parental income and racial backgrounds do not.

Overall, the results in this section are consistent with evidence from other domains that disparities in measurable skills are small at birth and expand gradually over time (e.g., Fryer and Levitt 2006, Fryer 2011). One explanation for these patterns is that differences in childhood environment – e.g., in the quality of schools or the degree of exposure to science and innovation – affect the amount students learn or the amount of time they study. However, as noted in prior work, one must be cautious in attributing these results to environmental differences. If tests at later ages are more effective at capturing intrinsic ability, one may find the patterns across grades documented above even in the absence of differences in childhood environment. In light of this limitation, we directly examine the causal effects of childhood environment in the next section.

IV Childhood Environment and Exposure to Innovation

In this section, we study how childhood environments affect innovation, focusing in particular on the role of exposure to inventors. We first exploit variation across technology classes to show that children's propensities to invent in a given field are heavily influenced by growing up with parents, parents' coworkers, or neighbors who are inventors. We then analyze the outcomes of children who move across areas to show that childhood environment affects not just the types of innovation that children pursue, but also the overall fraction who go into innovation.

IV.A Parents

To characterize the role that children's parents play in shaping their decision to pursue innovation, we begin by asking whether children whose fathers are inventors are more likely to become inventors themselves.[22] In our intergenerational analysis sample (children in the 1980-84 birth cohorts), 2.0

[22]We focus on fathers here because the vast majority of inventors, particularly in older generations, are male (Figure III). We examine the role of female inventors in the context of neighborhood differences, where we have greater power, in section IV.B below. We define a father as an inventor if he applied for a patent between 2001-2012 or was granted

out of 1,000 children whose parents were not inventors become inventors by 2014. In contrast, 18.0 per 1,000 children of inventors become inventors themselves – a nine-fold difference.[23] This pattern holds even conditional on parental income, across the parent income distribution (not reported).

The intergenerational persistence of innovation could be driven by the genetic transmission of ability to innovate across generations or by an *exposure effect* – the environmental effect of growing up in a family of innovators, holding one's intrinsic invention ability fixed. These exposure effects could reflect the accumulation of specific human capital, changes in preferences, or simply increased awareness about innovation as a career pathway.

We distinguish between intrinsic inventive ability and exposure effects by exploiting variation in the specific technology class in which a child innovates. Following the USPTO's classification system and Hall et al. (2001), patents can be grouped into seven broad categories (chemicals, computers and communications, drugs and medical, electrical and electronic, mechanical, design and plant, and other). Within these categories, patents are further classified into 37 sub-categories and 445 specific technology classes. These technology classes are very narrow: for instance, within the communications category, there are separate classes for modulators, demodulators, and oscillators; within the resins subcategory, there are separate classes for synthetic and natural resins.

We isolate the causal effects of exposure by analyzing whether children are particularly likely to patent in the same technology classes as their parents. The idea underlying our research design is that genetic differences in inventive ability are unlikely to lead to differences in propensities to innovate across similar, narrowly-defined technology classes. For instance, a child is unlikely to have a gene that codes specifically for ability to invent in modulators rather than oscillators. Under this assumption, the degree of alignment between the specific technology classes in which children and their parents innovate can be used to estimate causal exposure effects.

Implementing this research design requires a metric for the degree of similarity between technology classes. We define the distance between two technology classes A and B based on the share of inventors in class A who also invent in class B; the higher the share of common inventors, the lower the distance between A and B. Online Appendix Table III gives an example that illustrates this distance metric by showing the technology classes that are closest to technology class 375, "pulse

a patent between 1996-2014, analogous to the definition for children.

[23] Part of this association reflects the fact that children and their fathers sometimes are co-inventors on the same patent. However, this is relatively rare: 13.7 out of 1,000 children of inventors file patents on which their parent is not a co-inventor, still far higher than the rate for non-inventors. Additionally, our measure of parental inventor status suffers from measurement error because we do not observe parents' patents prior to 1996 in our data, likely attenuating our estimate of the difference.

or digital communications." Pulse or digital communications has a distance of zero with itself by definition. Inventors who had a patent in pulse or digital communications were most likely to have another patent in demodulators, which is therefore assigned an ordinal distance of $d = 1$ from the pulse and digital communications class. The next closest class is modulators ($d = 2$), and so on.

Figure VIIa plots the fraction of children who patent in a technology class d units away from their father's technology class, among children of inventors in our intergenerational sample.[24] Nearly 1 in 1,000 children patent in exactly the same technology class as their father ($d = 0$). In contrast, the probability of inventing in the next closest technology class (with distance $d = 1$) is less than 0.2 per 1,000, an estimate that is significantly different from the value at $d = 0$ with $p < 0.01$. The child's probability of inventing in a given class then falls gradually as d rises, although the gradient is relatively flat compared to the jump between $d = 0$ and $d = 1$.

The jump in innovation rates at $d = 0$ suggests that part of the reason that children of inventors are more likely to become inventors themselves is due to exposure to innovation rather than differences in natural talents. To formalize the identification assumption underlying this conclusion, let $e_{ic} \in \{0, 1\}$ represent an indicator for whether child $i's$ father has a patent in technology class c (i.e., if child i is "exposed" to innovation in class c) and a_{ic} represent the child's intrinsic ability to innovate in class c. Suppose that child i patents in technology class c if $a_{ic} + \beta e_{ic} > 0$. Here, β measures the causal effect of exposure to innovation. Our identification assumption is that:

$$\lim_{d \to 0} Cov(a_{i,c} - a_{i,c+d}, e_{i,c} - e_{i,c+d}) = 0. \tag{1}$$

Equation (1) requires that an individual's intrinsic ability to innovate in a technology class does not covary with whether his father innovates in that particular technology class among technology classes that are very similar. Under this assumption, we can identify the causal effects of exposure (β) even though inventive ability is correlated with exposure ($Cov(e_{ic}, \alpha_{ic}) > 0$) by analyzing how a child's propensity to innovate in a given technology class varies with the distance between that class and the class in which his parents patented. In particular, the jump in rates of innovation at $d = 0$ in Figure VII cannot be generated by differences in ability under the assumption in (1) and must therefore be driven by the causal effect of exposure.[25]

[24] Children or fathers who patent in multiple technology classes are assigned the technology class in which they patent most frequently. We omit observations where a child and his or her father are co-inventors on the same patent to eliminate mechanical effects on the rate of patenting in the same class.

[25] Equation (1) is a convenient way to conceptualize our research design, but we cannot literally take the limit as $d \to 0$ because of the discreteness of technology classes. In practice, we effectively assume that $Cov(\alpha_{i,c} - \alpha_{i,c+1}, e_{i,c} - e_{i,c+1}) = 0$, i.e. that a child's ability to invent in a technology class does not covary with parental exposure across two adjacent classes.

Interpreting the difference in innovation rates between technology class $d = 0$ and $d = 1$ as purely driven by exposure, we infer that having a parent who is an inventor in a given technology class increases a child's probability of inventing in that class by at least a factor of 5. This result suggests that exposure plays a substantial role in determining children's propensities to innovate.[26]

Although this result is useful in establishing that exposure matters, replicating the level of exposure one obtains through one's parents is likely to be challenging from a policy perspective. Moreover, parents are only one of many potential sources through which children may acquire knowledge about careers in innovation. We therefore turn to two broader sources of exposure outside one's immediate family: parents' coworkers and residential neighbors.

IV.B Parents' Coworkers

In this subsection, we examine how exposure to innovation through parents' coworkers affects a child's propensity to become an inventor. To do so, we first assign each father in our intergenerational sample an industry based on the six-digit NAICS code of his most frequent employer between 1999-2012.[27] We then measure the patent rate among workers in the father's industry – whom we term the father's "coworkers" – as the average number of patents issued to individuals in that industry per year (between 1996-2012) divided by the average number of workers in that industry per year based on counts of W-2 forms in the tax data. To ensure that we do not capture the effects of parental exposure itself, we drop children whose own parents were inventors during our sample period throughout the remainder of this section.[28]

In column 1 of Table III, we regress the fraction of children who become inventors among those with fathers in a given industry on patent rates for workers in that industry. This regression has one observation for each of the 345 industries and is weighted by the number of fathers in each industry.[29] The estimate of 0.250 (s.e. = 0.028) implies that a 1 percentage point increase in the patent rate among a father's coworkers is associated with a 0.25 percentage point increase in the

[26] More precisely, this research design demonstrates that parental exposure influences the technology class in which a child innovates. Although this finding supports the view that children whose parents are inventors are more likely to invent themselves because of exposure effects, one may be concerned that exposure affects only the *type* of innovation a child pursues and not whether or not the child invents at all. We address this possibility using an alternative research design in Section IV.D.

[27] For individuals receiving W-2s from multiple firms in a given year, we define the employer in that year to be the firm that issued the W-2 with the highest salary. We exclude fathers working in industries with fewer than 50,000 individuals (5% of fathers), as patent rates are measured imprecisely for these industries.

[28] To ensure that the findings are not driven by mechanical co-patenting with parents' co-workers, we have verified that restricting the sample to children who have sole-authored patents yields very similar results.

[29] This regression is equivalent to regressing an indicator for whether a child is an inventor on the rate of innovation in his or her father's industry in a dataset with one observation per child, clustering standard errors by industry, because the innovation rate (the right hand side variable) does not vary within industries.

probability that a child becomes an inventor. This estimate implies that a one standard deviation (0.24 pp) increase in the fraction of inventors in the father's industry is associated with a 25.3% (0.059 pp) increase in children's innovation rates.

The association in column 1 of Table III could reflect either the causal effect of exposure to innovation through a parents' coworkers or a correlation with other unobservables, such as a child's own intrinsic ability to innovate. As above, we isolate exposure effects by testing whether children are more likely to innovate in exactly the same technology classes as their parents' coworkers. Using the same measure of distance d between technology classes defined in Section IV.A, we estimate OLS regressions of the form:

$$y_{cj} = \kappa_c + b_d P_{c+d,j} + \varepsilon_{cj}, \tag{2}$$

where y_{cj} denotes the patent rate in technology class c of children with fathers who work in industry j, κ_c represents a class-specific intercept, and $P_{c+d,j}$ denotes the patent rate in the class $c + d$ among workers in industry j. We estimate these regressions at the industry by technology class level, weighting by the number of children with fathers in each industry. We include class fixed effects (κ_c) to account for the variation in size across classes and identify b_d from variation across industries in class-specific patent rates.

Figure VIIb plots estimates from regressions analogous to (2). Each bar plots estimates of b_d from a separate regression, varying the distance d used to define workers' patent rates $P_{c+d,j}$ in (2). The first bar plots b_0, the relationship between children's patent rates in a given class and their fathers' coworkers patent rates in the same class ($d = 0$). In the second bar, we define $P_{c+d,j}$ as the mean patent rate in the father's industry in the next 10 closest classes ($d = 1$ to 10). The third bar uses the average patent rate in classes with $d = 11$ to 20, and so on. The coefficient b_j on parents' coworkers' patent rates drops by 85% from the same class ($d = 0$) to the next closest classes ($p < 0.01$). That is, children are much more likely to patent in exactly the same class as their parents' coworkers than in very similar classes. This result implies that an increase in parents' coworkers patent rates *causes* an increase in a child's propensity to innovate under the following identification assumption:

$$\lim_{d \to 0} Cov(\varepsilon_{c,j} - \varepsilon_{c+d,j}, P_{c,j} - P_{c+d,j}) = 0. \tag{3}$$

This assumption, which is analogous to (1), requires that as the distance d between technology classes grows small, differences in unobservable determinants of children's innovation rates in class c vs. $c + d$ are orthogonal to differences in parents' coworkers' innovation rates in those classes.

Intuitively, we require that children whose fathers work in an industry where many workers patent in amplifiers rather than antennas do not have greater intrinsic ability to invent in amplifiers relative to antennas themselves. Under this assumption, we can infer from Figure VIIb that a 1 pp increase in patent rates among parental coworkers in a given class increases a child's probability of inventing in that class by $b_0 - b_{1-10} = 0.065$ pp (83%).

Our measure of distance between technology classes based on co-patenting rates is one of many potential approaches to identifying "similar" patent classes. To assess the sensitivity of our results to this choice, we use the Hall et al. (2001) hierarchical classification system, which groups patents into similar fields (categories, sub-categories, and classes), as an alternative way to identify similar patent classes. In columns 2-5 of Table III, we estimate a series of regressions to assess whether children patent in the same fields as workers in their father's industry using the USPTO's classification system. In column 2, we test whether children are more likely to invent in the same categories as their father's coworkers using a regression specification analogous to (2) estimated at the category by industry level with $d = 0$. Columns 3 and 4 replicate the specification in column 2 at the sub-category and technology class levels. Finally, in column 5 of Table III, we replicate column 4 with three additional controls: patent rates in (i) the same sub-category but in a different patent class, (ii) the same category but a different sub-category, and (iii) other categories.

At all levels of the hierarchy, we find a strong, statistically significant association between children's patent rates and their parents' coworkers patent rates. Moreover, column 5 shows that innovation among parents' coworkers leads to a 10 times larger increase in innovation in exactly the same technology class (e.g., synthetic resins) as it does in other classes even within the same sub-category (e.g., natural resins). The coefficient on the own-class patent rate is not statistically different from the specification in column 4, while the coefficients on the other-class and category patent rates are very close to zero. Under our identification assumption in (3), the much smaller estimates for other classes imply that children's propensity to invent in the same class as their parents' coworkers is driven by the causal effect of exposure.

The class-specificity of the exposure effects also sheds light on the mechanism through which exposure matters. Transmission of general human capital or an interest in science would be unlikely to have impacts that vary so sharply by technology class. Instead, the data point to mechanisms such as transmission of specific human capital, access to networks that help children pursue a certain subfield, acquisition of information about certain careers, or role model effects.

IV.C Neighborhoods

In this subsection, we study how rates of innovation in the neighborhood in which a child grows up affect his or her propensity to innovate. Following Chetty et al. (2014b), we assign children in our intergenerational sample to commuting zones (CZs) based on where they were first claimed as dependents by their parents.

Figure VIIIa maps rates of innovation across the CZs where children grew up, with darker colors representing areas where more children become inventors. Figure VIIIb lists the ten CZs where children are the most or least likely to grow up to become inventors (among the 100 most populated CZs). Children who grow up in the Northeast, coastal California, and the rural Midwest have the highest probabilities of becoming inventors, while those in the Southeast have the lowest probability. The areas where children grow up to become inventors tend to have higher mean incomes (population-weighted correlation $\rho = 0.63$), fewer single parents ($\rho = -0.39$), and higher levels of absolute upward intergenerational mobility ($\rho = 0.32$), based on the CZ-level measures defined in Chetty et al. (2014b). However, there are some stark exceptions to these patterns, such as Detroit, MI, where children have among the highest likelihood of becoming inventors but where income mobility and mean incomes are relatively low.

The spatial analysis in Figure VIII differs from previous analyses of "innovation clusters" and agglomeration (e.g., Porter and Stern 2001, Kim and Marschke 2005) because it reflects the locations where inventors grow up, which may differ from where they work as adults. Nevertheless, children who grow up in the areas where the most innovation occurs tend to be most likely to go into innovation themselves. For instance, children who grow up in the San Jose commuting zone, which includes Silicon Valley, top the list in terms of the probability of becoming inventors themselves. To examine this relationship more systematically, we define the patent rate of workers in each CZ as the average number of patents issued per year (in the full USPTO data) to individuals from a given CZ between 1980 and 1990 divided by the CZ's population between the ages of 15-64 in the 1990 Census. Figure IX presents a scatter plot of the fraction of children who go on to become inventors vs. the patent rate of workers in their childhood CZ (their "neighbors") among the 100 most populated CZs. There is a clear positive relationship between these variables, with a correlation of 0.75.

The correlation in Figure IX is consistent with the hypothesis that exposure to innovation during childhood through one's neighbors increases a child's propensity to innovate, but it could also reflect geographical sorting. We isolate the causal effect of exposure by estimating the extent to which children invent in the same narrow technology classes as their neighbors, as in our analysis of industry-level differences above. Figure VIIc replicates Figure VIIb, plotting coefficients from regressions of children's innovation rates in a given technology class c on class-level patent rates of workers in their childhood CZs vs. the distance between technology classes. The coefficient on neighbors' patent rates drops by 85% from the same class ($d = 0$) to the next closest classes ($p < 0.01$), implying that neighborhoods have substantial causal exposure effects on the class in which a child innovates under an identification assumption analogous to (3).

In Table IV, we evaluate the robustness of this result and the mechanisms underlying it using a set of fixed effects regression specifications. As a reference, in column 1, we regress the fraction of children who grow up to be inventors in each CZ on the patent rate of workers in their childhood CZ, replicating the analysis in Figure IX including all 741 CZs rather than just the 100 largest ones. The coefficient of 2.9 implies that a 1 SD (0.02 pp) increase in the annual CZ-level patent rate is associated with a 0.058 pp (28.5%) increase in the fraction of children who become inventors.

One potential explanation for the result in column 1 (and Figure VIIc) is that children tend to stay near the areas where they grew up, and may mechanically end up being more likely to patent if they live in an area like Silicon Valley simply because the jobs that are available in such areas tend to be in the innovation sector. To distinguish this mechanism from childhood exposure effects that change children's behavior, we focus on the subset of children who move to a different CZ in adulthood from where they grew up. In column 2, we estimate a regression analogous to that in column 1 at the childhood CZ by current CZ level, limiting the sample to children whose current (2012) CZ differs from their childhood CZ. We regress the fraction of children who grow up to be inventors in each of these cells on the patent rate of the CZ in which they grew up, including fixed effects for the child's 2012 CZ so that the coefficient of interest is identified purely from comparisons across individuals who grew up in different areas but currently live in the same area. The coefficient on the patent rate in the childhood CZ is only slightly lower at 2.6 in this specification (compared to 2.9 in column 1), showing that most of the relationship in column 1 is not mechanically driven by the types of jobs available in an area.

In the remaining columns of Table IV, we use the hierarchical patent classification system to identify similar patent classes instead of the distance metric used in Figure VII. In columns 3-5, we analyze whether the result in column 1 continues to hold at the category level: do children go on to patent in the same categories as their neighbors did while they were growing up? We consider three different specifications. In column 3, we replicate the specification in column 1 at the CZ by patent category level, using the same specification as in (2) when $d = 0$, but letting j index CZs instead of industries. In column 4, we replicate the specification in column 2 at the category level. We restrict attention to movers and regress the share patenting in a given category (with one observation per childhood CZ, current CZ, and category) on the childhood CZ patent rate in that category. We include current CZ by category fixed effects in this specification. In column 5, we include all children and replace the CZ by category fixed effects with fixed effects for the father's industry by category, estimating the model at the childhood CZ by father's industry by category level. This specification isolates variation from one's neighbors that is orthogonal to the variation from parents' coworkers examined above in Table III.

In all three of these specifications in Table IV, we find robust and significant positive relationships between children's category-level innovation rates and the corresponding category-level patent rates of workers in their childhood CZ. Intuitively, these specifications effectively show that children who grow up in Silicon Valley are especially likely to patent in computers, while children who grow up in Minneapolis (which has many medical device manufacturers) are especially likely to patent in medical devices. This is true even among children who live in the same place in adulthood and whose parents work in the same industry.

In columns 6 and 7 of Table IV, we replicate the specification in column 3 at the sub-category and technology class levels, respectively. We continue to find substantial positive coefficients in these specifications, confirming the result in Figure VIIc that children tend to invent in the same technology classes that those around them did during their childhood. Column 8 replicates the specification in column 7 including controls for patent rates in other classes, sub-categories, and categories, as in column 5 of Table III. The coefficient on the own-class coefficient is not statistically different from the specification in column 7, while the coefficients on the other-class and category patent rates are close to zero. Under our identification assumption, the coefficient of 1.02 in column 8 implies that a 1 SD (0.0002 pp) increase in the annual CZ-level patent rate in a given technology class causes a 0.0002 pp (43%) increase in the fraction of children who become inventors in the same class.

Gender-Specific Exposure Effects. Next, we examine the heterogeneity of exposure effects by gender, focusing specifically on whether girls are more likely to go into innovation if they are exposed to female inventors as children. As a first step, Figure X shows how gender gaps in innovation vary across the areas in which children grow up using our intergenerational analysis sample. Panel A maps the fraction of female inventors by the state in which inventors grew up, while Panel B shows this statistic for the top 10 and bottom 10 CZs among the 100 largest CZs.[30] Although no state comes close to gender parity, there is significant variation in the magnitude of the gender gap: 28.7% of children who grow up to become inventors in Rhode Island are female, as compared with 11.3% in Idaho.[31]

To test whether gender-specific differences in exposure to innovation lead to the differences in gender gaps in Figure X, we first estimate gender-specific patent rates for workers in each CZ. We do so using our linked patent-tax sample instead of all patents in the USPTO data as above because gender is not observed in the USPTO data.[32] As a benchmark, column 1 of Table V replicates the specification in column 1 of Table IV using this alternative measure of the CZ-level innovation rate. The raw magnitude of the coefficient differs because the tax-data-based innovation rate is scaled differently from the USPTO-based measure. However, a 1 SD increase in the CZ innovation rate is associated with a 30.8% increase in children's propensities to innovate, very similar to the 28.5% estimate obtained above in column 1 of Table IV.

In column 2 of Table V, we regress the fraction of females who go on to patent in each CZ on the innovation rates for women and men in that CZ. The coefficient on female innovation rate is significant and positive, while the coefficient on the male innovation rate is small and statistically insignificant. Symmetrically, column 3 shows that male innovation rates are more predictive of boys' propensities to become inventors than female innovation rates.[33] These estimates imply that if girls were as exposed to female inventors in their childhood CZs as boys are to male inventors, female innovation rates would rise by 164% and the gender gap in innovation would fall by 55%.[34]

[30] We present this map at the state level because gender-specific patent rates are noisy in small CZs due to the small number of female inventors.

[31] The gender gap is generally smaller in states that score higher on Pope and Sydnor (2010)'s gender stereotype adherence index on standardized tests in 8th grade, which measures the extent to which children in a state adhere to the stereotype that boys are better at math and science while girls are better at English (population-weighted correlation = 0.21; Online Appendix Figure IV).

[32] Specifically, we define the innovation rate for gender g in CZ j as the total number of patent applications filed by individuals of gender g born before 1980 in our full inventors sample divided by the number of individuals between ages 15 and 64 of gender g in CZ j in the 1990 Census. We convert this measure to an annual rate by dividing by 17, as we observe patent applications between 1996-2012. We restrict attention to inventors born before 1980 to avoid overlap with the intergenerational analysis sample that we use to study outcomes. Pooling genders, the populated-weighted correlation across CZs between this measure of innovation rates and the USPTO-based measure used above is 0.65.

[33] We find similar patterns at the individual level – daughters are more likely to become inventors if their mothers are inventors while sons are more likely to become inventors if their fathers are inventors – but the coefficients are imprecisely estimated because there are so few female inventors among parents in our intergenerational sample.

[34] We estimate the counterfactual innovation rate for girls by adding to the current innovation rate for girls the difference in exposure to own-sex inventors for boys versus girls multiplied by the coefficient of 2.408 in column 2. To calculate the difference in the gender gap, we similarly use the estimates of the effect of exposure to adult female inventors on *both* boys and girls (columns 2 and 3) to predict how the patenting rates of both genders would change if

One potential concern with the analysis in columns 2 and 3 is that women may have particularly strong tastes or abilities to innovate in certain fields (e.g., biology). This could generate the gender-specific associations in columns 2 and 3 even in the absence of exposure effects if children live in the same areas as their parents and the types of jobs (e.g., biology vs. information technology) varies across places. Columns 4 and 5 evaluate this concern by examining variation in innovation rates across patent categories, using a specification with one observation per CZ by patent category with category fixed effects, as in column 3 of Table IV. We find very similar patterns in these specifications: women are more likely to innovate in a particular category if there were more women innovating in that category in the area where they grew up. We reject the null hypothesis that the coefficients are the same for both genders with $p < 0.02$ in both of these specifications, implying that the findings in columns 2 and 3 are not due to selection across categories.

In sum, Table V further supports the hypothesis that exposure to innovation in childhood through one's neighbors has a causal effect on children's propensities to pursue innovation by providing an additional overidentification test of that hypothesis. In particular, the results in Table V imply that any confounding variable would have to vary not just across technology classes, but also in a gender-specific manner. Moreover, these findings suggest that the differences in rates of innovation across areas where children grow up are unlikely to be driven purely by factors such as schools or segregation emphasized in prior work on neighborhood effects, as such factors would be unlikely to generate impacts that vary so sharply by gender and technology class.

IV.D Neighborhood Effects on the Level of Innovation

The technology class-level results in the preceding subsections show that exposure affects the type of innovation one pursues, but they do not necessarily imply exposure matters for whether one chooses to become an inventor to begin with. In this section, we examine whether exposure also affects the *level* of innovation. To do so, we study how the patent rates of children who move across areas vary with the age at which they move. Chetty and Hendren (2018) use this timing-of-move design to establish that neighborhoods have causal effects on children's earnings. Here, we use the same design to study the impacts of neighborhoods on the fraction of children who patent in adulthood. Intuitively, we ask: "Are children who move to high-innovation areas at younger ages

exposure to female inventors were as high as it is to male inventors. Naturally, these estimates should be interpreted with caution as they rely on out-of-sample linear extrapolations. We defer quantification of the extent to which exposure explains gaps in innovation by parental income and race to future work, as we lack analogous measures of exposure along these dimensions because we only observe race in the New York City school district sample and there are very few inventors who come from low-income families.

more likely to become inventors (in any field) themselves?" Under the assumption that the children who make a given move at earlier vs. later ages are comparable to each other, the answer to this question reveals the extent to which neighborhoods have causal effects on children's propensities to invent.[35]

Empirical Specification. We study the outcomes of children who move across CZs exactly once during their childhood using our intergenerational sample, which we extend to include birth cohorts 1980-88 in order to expand the range of ages at move we can observe. Let i index children. In the sample of one-time movers, let m_i denote the age at which child i moves from origin CZ o to destination CZ d. Chetty and Hendren (2018) show that neighborhoods have causal exposure effects on earnings and a variety of other outcomes before age 24; we therefore focus on moves that occur at or before age 24 in our analysis.[36]

As in the previous subsection, we define the patent rate among adults in each CZ as the average number of patents issued per year (in the full USPTO data) to individuals from a given CZ between 1980 and 1990, divided by the CZ's population between the ages of 15-64 in the 1990 Census. Let $\bar{p}_d$ and $\bar{p}_o$ denote the patent rates in the destination and origin CZs and $\Delta_{od} = \bar{p}_d - \bar{p}_o$ denote the difference in patent rates in the destination versus origin CZ.

After computing these variables, we regress an indicator for whether the child becomes an inventor by 2012 (y_i) on the measures of origin and destination patent rates interacted with the child's age at move:

$$y_i = a + \beta m_i \Delta_{od} + \gamma_1 \Delta_{od} + \gamma_2 X_i + \epsilon_i \tag{4}$$

where X_i denotes a control vector that includes age at move fixed effects, birth cohort fixed effects, and other controls that we vary across specifications. The key parameter of interest is β, which captures how a child's propensity to become an inventor varies with the age at which he or she moves to an area with higher patent rates.

Identification Assumptions. We can interpret β as the causal effect of one additional year of exposure to a higher-innovation area (i.e., an area with higher observed patent rates) during childhood, under the assumption that the potential outcomes of children who move to better vs. worse

[35] Critically, this research design does not require that *where* people move is orthogonal to their potential outcomes; it simply requires that the *timing* of those moves is unrelated to potential outcomes.

[36] More precisely, Chetty and Hendren (2018, Figure IV) demonstrate that children's earnings (and other outcomes) decline linearly with age at move (m) up to age 24 and are constant thereafter. Motivated by this functional form, we include moves that occur after age 24 by defining $m_i = 24$ for such moves in order to maximize to maximize the precision of our estimates. Excluding moves above age 24 yields qualitatively similar but less precise estimates.

areas do not vary with the age at which they move. Chetty and Hendren (2018) present a series of tests supporting this orthogonality condition: controlling for unobserved heterogeneity across families using sibling comparisons in models with family fixed effects, implementing a set of placebo tests exploiting heterogeneity in predicted causal effects across subgroups, and validating the results using experimental designs, e.g. from the Moving to Opportunity Experiment (Chetty et al. 2016). They also show that the relationship between children's outcomes and age at move declines linearly up to age 23 and is flat thereafter, justifying the linear specification in (4). Furthermore, Chetty and Hendren (2018) provide evidence that estimates of place effects among movers are externally valid to the broader population because they find similar results among those who self-select to move as compared to families displaced by idiosyncratic events such as hurricanes. Building on these results, we take the validity of the research design and empirical specification in (4) as given here and apply it to identify the causal effects of neighborhoods on patent rates.[37]

Results. Table VI reports estimates of β for several variants of (4). In column 1, we estimate (4) including origin fixed effects, effectively comparing children who start in the same CZ but move to different CZs. We obtain an estimate of $\beta = -0.08$ ($p < 0.01$). This estimate implies that if a child grows up for 20 years in a CZ with a patent rate among adults that is 1 SD (0.02 pp) above the mean, then his likelihood of becoming an inventor increases by $20 \times 0.08 \times 0.02 = 0.032$ percentage points (22%).

Columns 2 and 3 of Table VI present variants of the specification in Column 1 to assess the robustness of the estimates. In column 2, we control for the origin patent rate instead of including origin fixed effects. This more parsimonious specification yields a very similar estimate of $\beta = -0.08$. In column 3, we include interactions of the change in patent rates and the origin patent rate with indicators for the child's birth cohort to account for the fact that children's propensities to invent by 2012 will naturally vary across cohorts. This specification again yields quite similar estimates.

To gauge the magnitude of these exposure effect estimates, in column 4 of Table VI we report estimates from a cross-sectional regression of an indicator for whether a child invents on the patent rate of adults in the first CZ in which we observe the child living, including both movers and non-movers. This specification replicates the cross-sectional regression presented above in column 1 of

[37]Since patenting is a relatively rare outcome, we lack the precision to replicate the non-parametric specifications and additional tests implemented by Chetty and Hendren (2018); for instance, specifications that include family fixed effects yield point estimates similar to our baseline estimates but are statistically insignificant. However, given that Chetty and Hendren (2018) establish the validity of the design for several outcomes that are highly correlated with innovation, such as earnings and college attendance, we believe the design is likely to be valid for patenting as well.

Table IV using the extended set of birth cohorts (1980-88) that we use in our movers analysis. The coefficient of 2.04 in Column 4 implies that a 1 pp increase in the annual patent rate among adults in a CZ is associated with a 2.04 pp increase in the fraction of children who become inventors. Under our identification assumptions, Columns 1-3 imply the causal effect of growing up in a neighborhood (for 20 years of childhood) with 1 pp higher patent rates among adults increases children's patent rates in adulthood by $20 \times 0.08 = 1.6$ pp. Hence, approximately 75% (˜1.6/2) of the cross-sectional relationship between innovation rates and children's probability of inventing documented above in Figure IX is due to neighborhood-level exposure effects on the level of innovation.

The estimates above imply that moving a child from a CZ that is at the 25th percentile of the distribution in terms of inventors per capita (e.g., New Orleans, LA) to the 75th percentile (e.g., Austin, TX) – a 1.4 SD change – would increase his or her probability of becoming an inventor by $1.4 \times .032 = .045$ percentage points (37%). Hence, exposure to innovation has substantial impacts not just on the types of innovation children pursue but also on whether or not they become inventors at all.

V Inventors' Careers: The Potential for Lost Einsteins

Are the children from low-income families who do not pursue innovation (e.g., because of a lack of exposure) individuals who would have ended up having highly impactful innovations? Or do the most productive "stars" overcome the hurdles they face and become inventors regardless of their background, as predicted by economic models of career selection with barriers to entry (Hsieh et al. 2016)? In this section, we adress this question by analyzing how the returns to innovation vary with inventors' characteristics at birth.

We consider two measures of returns to innovation: inventors' earnings (a measure of private returns) and patent citations (a proxy for social impact). As reference, we plot the income distribution of inventors between ages 40-50 in our sample in Online Appendix Figure Va. The distribution is extremely skewed: the median annual income (in 2012 dollars) is \$114,000, the mean is \$192,000, and the 99th percentile is \$1.6 million. The private returns to innovation are highly correlated with their social impact, as measured by citations (Online Appendix Figure Vb). Notably, inventors who have patents in the top 1% of the citation distribution earn more than \$1 million per year between ages 40 and 50, confirming that highly-cited patents are highly valued by the market.

If the factors that lead to an under-representation of inventors from certain groups (e.g., low-income families) screen out inventors who would have had low returns, inventors from those groups will have *higher* returns on average than inventors from more advantaged backgrounds (Hsieh et al. 2016).[38] We test whether this is the case in Figure XI. In Panel A, we compare the mean incomes of inventors with different characteristics at birth. The first pair of bars compares individuals from families with incomes above vs. below the 80th percentile of the parental income distribution using inventors in our intergenerational analysis sample. The second pair compares minorities (Blacks and Hispanics) to non-minorities using inventors in the New York City schools sample. The third pair compares males and females using the full inventors sample. In all cases, inventors from the under-represented groups have similar or lower earnings on average than those from more advantaged backgrounds.

Figure XIb replicates this analysis using the probability of having a highly-cited patent (in the top 5% of the distribution of citations among inventors in a given birth cohort) as the outcome. The patterns are analogous: inventors from under-represented groups also do not have higher-impact inventions.

Figure XI implies that the probability that an individual becomes a star (high-return) inventor is just as sensitive to his or her conditions at birth as the probability that he innovates at all, as shown in Figure Ib in the context of parental income. This finding is consistent with our conclusion above that differences in exposure to innovation play a key role in generating these gaps. A lack of exposure (e.g., awareness of innovation as a potential career) is likely to reduce the probability that individuals pursue innovation uniformly across all levels of productivity. In contrast, this result challenges standard economic models that explain differences in occupational choice purely by differences in barriers to entry across subgroups (e.g., Hsieh et al. 2016), because such models predict that the marginal inventors who are screened out are those with lower potential. In order to explain the patterns, the factors that generate barriers to entry must also reduce individuals' productivity after entering innovation (e.g., discrimination).[39]

Regardless of the explanation, the key implication of Figure XI is that there are many "lost Einsteins" – individuals who do not pursue a career in innovation even though they would have had highly impactful innovations had they done so. To quantify the amount of lost innovation, we consider a counterfactual under which women, minorities, and children from low-income (bottom 80%) families invent at the same rate as white men from high-income (top 20%) families.[40] In this

[38]This result assumes that the ability to innovate does not vary across groups – an assumption supported by the evidence in Section III.B.

[39]Indeed, the fact that inventors who are women, minorities, or from lower-income families are all paid less than their more advantaged counterparts despite having similar citations is consistent with on-the-job discrimination.

[40]Since we do not observe patent rates in the full population by race, we implement this calculation by assigning

scenario, there would be 4.04 times as many inventors in America as there are today.[41] Although one cannot conclude that aggregate welfare would be higher if these "lost Einsteins" were to enter innovation rather than the careers they currently pursue, this calculation does illustrate that focusing on the "extensive margin" of the supply of inventors is valuable if one's objective is to increase aggregate innovation.

In the working paper (Bell et al. 2017), we present a more comprehensive analysis of inventors' careers trajectories and compare the effects of alternative policies to increase innovation using a stylized model of career choice that incorporates exposure effects. The model implies that the potential to increase innovation by increasing financial returns (e.g., by cutting top income tax rates) is limited because such policies only affect the subset of individuals who have exposure and because the decisions of star inventors – who earn very high salaries already – are unlikely to be affected by marginal changes in incentives (Jaimovich and Rebelo 2017). In contrast, increasing exposure can have substantial impacts on quality-weighted innovation by drawing individuals who produce high-impact inventions into the innovation pipeline. We therefore conclude that there is substantial scope to increase aggregate innovation by increasing exposure to innovation among under-represented groups.

VI Conclusion

This paper has presented new evidence on the factors that determine who becomes an inventor by tracking the lives of inventors in America from birth to adulthood. Most previous work on innovation has focused on factors such as financial incentives, barriers to entry, and STEM education. Our results point to a different channel – exposure to innovation during childhood – as a critical factor that determines who becomes an inventor and the types of innovations they pursue. A lack of exposure to innovation can help explain why talented children in low-income families, minorities, and women are significantly less likely to become inventors. Importantly, such lack of exposure potentially screens out not just marginal inventors but the "Einsteins" who produce innovations that have the greatest impacts on society. Policies that increase exposure therefore have the capacity to greatly increase quality-weighted aggregate innovation.

Policies to increase exposure to innovation could range from mentoring by current inventors

all individuals the patent rates of men born to parents in the top quintile (pooling races). In practice, data from the 2000 Census show that 86% of individuals in the top household income quintile are white, so this turns out to be a good approximation of patent rates for high-income white men.

[41] Of course, this calculation does not account for general equilibrium effects: such a large increase in the number of inventors might reduce the returns to innovation both privately and socially.

to internship programs at local companies. Our analysis does not provide guidance on which specific programs are most effective, but it does provide some guidance on how they should be targeted. In particular, targeting exposure programs to women, minorities, and children from low-income families who excel in math and science at early ages (e.g., as measured by performance on standardized tests) is likely to maximize their impacts on innovation. Furthermore, tailoring programs to participants' backgrounds may increase their impact; for example, our findings suggest that women are more influenced by female inventors rather than male inventors.

Beyond the literature on innovation, our findings contribute to the growing literature on how children's prospects for success are shaped by their environments. Prior studies have focused primarily on general human capital accumulation as the mechanism through which neighborhoods and schools affect outcomes. Our analysis suggests that environment matters through much narrower channels as well, for instance by influencing the specific career pathways that children choose to pursue, either via transmission of specific human capital or through changes in aspirations. Such mechanisms call for a different class of interventions than traditional investments in schools or neighborhoods, such as programs or networks that provide children exposure to specific careers that may be a good match for their talents.

More broadly, our findings suggest that policies designed to increase intergenerational mobility may also be beneficial for increasing economic growth. Drawing more low-income and minority children into science and innovation could increase their incomes – thereby reducing the persistence of inequality across generations – while stimulating growth by harnessing currently under-utilized talent. 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. Developing and testing methods to increase exposure to innovation among disadvantaged subgroups is therefore a particularly promising direction for research and policy.

References

Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen (2017). The Social Origins of Inventors. *mimeo*.

Aghion, P. and P. Howitt (1992). A Model of Growth Through Creative Destruction. *Econometrica 60*(2), 323–351.

Akcigit, U., J. Grigsby, and T. Nicholas (2017). The Birth of American Ingenuity: Innovation and Invention in the Golden Age. *NBER Working Paper No. 23047*.

Azoulay, P., J. Graff Zivin, and G. Manso (2011). Incentives and Creativity: Evidence from the Academic Life Sciences. *RAND Journal of Economics 42*(3), 527–554.

Becker, B. (2015). Public R&D Policies and Private R&D Investment: A Survey of the Empirical Evidence. *Journal of Economic Surveys 29*(5), 917–942.

Bell, A. M., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2017, November). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. Working Paper 24062, National Bureau of Economic Research.

Bloom, N., M. Schankerman, and J. Van Reenen (2013). Technology Spillovers and Product Market Rivalry. *Econometrica 81*(4), 1347–1393.

Card, D. and L. Giulano (2014). Does Gifted Education Work? For Which Students? *NBER Working Paper No. 20453*.

Chetty, R., J. Friedman, E. Saez, N. Turner, and D. Yagan (2017). Mobility Report Cards: The Role of Colleges in Intergenerational Mobility. *NBER Working Paper No. 23618*.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How Does your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R. and N. Hendren (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R., N. Hendren, and L. Katz (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review 106*(4), 855–902.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014b). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Cook, L. and C. Kongcharoen (2010). The Idea Gap in Pink and Black. *NBER Working Paper No. 16331*.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica 64*(5), 1001–1044.

Ding, W., F. Murray, and T. Stuart (2006). Gender Differences in Patenting in the Academic Life Sciences. *Science 313*(5787), 665–667.

Dorner, M., D. Harhoff, and K. Hoisl (2014). Patterns and Determinants of Inventor Mobility – Evidence from the Employment Biographies of Inventors in Germany. *mimeo*.

Fryer, R. and S. Levitt (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics 86*(2), 447–464.

Fryer, R. and S. Levitt (2006). The Black-White Test Score Gap through Third Grade. *American Law and Economics Review 8*(2), 249–281.

Fryer, R. G. (2011). Racial inequality in the 21st century: The declining significance of discrimination. In *Handbook of labor economics*, Volume 4, pp. 855–971. Elsevier.

Giannetti, M. and A. Simonov (2009). Social interactions and entrepreneurial activity. *Journal of Economics & Management Strategy 18*(3), 665–709.

Giuri, P., M. Mariani, S. Brusoni, G. Crespi, D. Francoz, A. Gambardella, W. Garcia-Fontes, and A. Geuna (2007). Inventors and Invention Processes in Europe: Results from the PatVal-EU Survey. *Research Policy 36*(8), 1107–1127.

Goolsbee, A. (1998). Does Government R&D Policy Mainly Benefit Scientists and Engineers? *American Economic Review 88*(2), 298–302.

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature 28*(4), 1661–1707.

Hall, B., A. Jaffe, and M. Trajtenberg (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. *NBER Working Paper No. 8498*.

Hall, D. T. (2002). *Careers In and Out of Organizations*. Sage Publications.

Heckman, J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science 312*(5782), 1900–1902.

Hsieh, C., E. Hurst, C. Jones, and P. Klenow (2016). The Allocation of Talent and U.S Economic Growth. *NBER Working Paper No.18693*.

Hunt, J. (2009). Why Don't Women Patent? *NBER Working Paper No. 17888*.

Jaimovich, N. and S. Rebelo (2017). Non-linear Effects of Taxation on Growth. *Journal of Political Economy 125*(1), 265–291.

Jones, B. (2010). Age and Great Invention. *The Review of Economics and Statistics 92*(1), 1–14.

Jones, C. and J. Williams (1999). Too Much of a Good Thing? *Journal of Economic Growth 5*(1), 65–85.

Jung, T. and O. Ejermo (2014). Demographic Patterns and Trends in Patenting: Gender, Age and Education of Inventors. *Technological Forecasting and Social Change 86*, 110–124.

Kahn, S. and D. Ginther (2017). Women and STEM. *NBER Working Paper No. 19894*.

Khan, B. Z. and K. L. Sokoloff (1993). schemes of practical utility: entrepreneurship and innovation among great inventors in the united states, 1790–1865. *The Journal of Economic History 53*(2), 289–307.

Kim, J. and G. Marschke (2005). Labor Mobility of Scientists, Technological Diffusion, and the Firm's Patenting Decision. *RAND Journal of Economics 36*(2), 298–317.

Laband, D. and B. Lentz (1983). Like Father, like Son: Toward an Economic Theory of Occupational Following. *Southern Economic Journal 50*(2), 474–493.

Lentz, B. and D. Laband (1989). Why So Many Children of Doctors Become Doctors: Nepotism vs. Human Capital Transfers. *Journal of Human Resources 24*(3), 396–413.

Lindquist, M., J. Sol, and M. Van Praag (2015). Why Do Entrepreneurial Parents Have Entrepreneurial Children? *Journal of Labor Economics 33*(2), 269–296.

Murphy, K., A. Shleifer, and R. Vishny (1991). The Allocation of Talent: Implications for Growth. *The Quarterly Journal of Economics 106*(2), 503–530.

Nanda, R. and J. B. Sørensen (2010). Workplace peers and entrepreneurship. *Management Science 56*(7), 1116–1126.

Nicholas, T. (2010). The role of independent invention in us technological development, 1880–1930. *The Journal of Economic History 70*(1), 57–82.

Nicolaou, N., S. Shane, L. Cherkas, J. Hunkin, and T. D. Spector (2008). Is the tendency to engage in entrepreneurship genetic? *Management Science 54*(1), 167–179.

Organisation for Economic Co-operation and Development (2009). *OECD Patent Statistics Manual 2009*. OECD.

Pope, D. and J. Sydnor (2010). Geographic Variation in the Gender Differences in Test Scores. *Journal of Economic Perspectives 24*(2), 95–108.

Porter, M. E. and S. Stern (2001). Innovation: Location Matters. *MIT Sloan Management Review 42*(4), 28–36.

Romer, P. (2000). Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers? *NBER Working Paper No. 7723*.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of political Economy 98*(5, Part 2), S71–S102.

Schmookler, J. (1957). Inventors Past and Present. *The Review of Economics and Statistics 39*(3), 321–333.

Shane, S. and N. Nicolaou (2013). The genetics of entrepreneurial performance. *International Small Business Journal 31*(5), 473–495.

Strumsky, D. (2014). Strumsky Patent Database (Available upon Request). Downloaded in December 2014.

Thursby, M. and J. Thursby (2005). Gender Patterns of Research and Licensing Activity of Science and Engineering Faculty. *Journal of Technology Transfer 30*(4), 343–353.

Toivanen, O. and L. Vaananen (2012). Returns to Inventors. *The Review of Economics and Statistics 94*(4), 1173–1190.

Topel, R. and M. Ward (1992). Job Mobility and the Careers of Young Men. *Quarterly Journal of Economics 107*(2), 439–479.

ONLINE APPENDIX

Matching Algorithm

The patent data were linked to the tax records using a variant of the matching algorithm developed in Chetty et al. (2014a) to link the New York City school district data to the tax records. Chetty et al. (2011) show that the match algorithm outlined below yields accurate matches for approximately 99% of cases in a school district sample that can be exactly matched on social security number. Note that identifiers were used solely for the matching procedure. After the match was completed, the data were de-identified (i.e., individual identifiers such as names were stripped) and the statistical analysis was conducted using the de-identified dataset.

Before beginning the match process, the names were standardized as follows. First, suffixes sometimes appear at the end of taxpayers' first, middle, or last name fields. If these fields end with a space followed by "JR", "SR", or a numeral I-IV, the suffix is stripped out and stored separately from the name. Second, the USPTO database separates inventor names into "first" and "last," but the tax data often separates names into first, middle, and last. In practice, many inventors include a middle initial or name in the first name field. Whenever there is a single space in the inventor's first name field, for the purposes of matching, we allow the first string to be an imputed first name, and the second string to be an imputed middle name or initial. The use of these imputed names is described below.

The matching algorithm proceeds in seven steps. Inventors enter a match round only if they have not already been matched to a taxpayer in an earlier round. Each round consists of a name criterion and a location criterion. The share of data matched in each round is documented below.

- **Stage 1**: Exact match on name and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name. Either the inventor's first name field exactly matches the concatenation of the IRS first and middle name fields or the IRS middle name field is missing, but the first name fields match. If an imputed middle name is available for the inventor, candidate matches are removed if they have ever filed at the IRS with a middle name or initial that conflicts with the inventor's.
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 49% of patents are uniquely matched in this stage.

- **Stage 2**: Exact match on imputed name data and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name and the taxpayer's last name is the same as the inventor's imputed first name. Either the inventor's imputed middle name/initial matches one of the taxpayer's middle/initial name fields, or one of the two is missing. For inventors with non-missing imputed middle names, priority is given to matches to correct taxpayer middle names rather than to taxpayers with missing middle names. As above, candidate matches are removed if they have ever filed at the IRS with a conflicting middle name or initial.
  - Location match: As above, the inventor's city and state must exactly match some city and state reported by that taxpayer.
  - 12% of patents are uniquely matched in this stage.

- **Stage 3**: Exact match on actual or imputed name data and 1040 zip crosswalked.

  - Name match: The inventor's last name exactly matches the taxpayer's last name. The inventor's first name matches the taxpayer's first name in one of the following situations, in order of priority: (1) inventor's first name is the same as the taxpayer's combined first and middle name; (2) inventor's imputed first name matches taxpayer's and middle names match on initials; (3) inventor has no middle name data, but inventor's first name is the same as the taxpayer's middle name.
  - Taxpayers are removed if they are ever observed filing with middle names in conflict with the inventor's.Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 3% of patents are uniquely matched in this stage.

- **Stage 4**: Same as previous stage, but using names from 1040 forms instead of names from W-2 forms.
  - Name match: The inventor's name matches the name of a 1040 (or matches without inventor's middle initial/name and no taxpayer middle initials/names conflict with inventor's).
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 6% of patents are uniquely matched in this stage.

- **Stage 5**: Match using W-2 full name field.
  - Name match: The inventor's FULL name exactly matches the FULL name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 8% of patents are uniquely matched in this stage.

- **Stage 6**: Fuzzy match using W-2 full name field.
  - Name match: The inventor's full name (minus the imputed middle name) exactly matches the full name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 1% of patents are uniquely matched in this stage.

- **Stage 7**: Match to all information returns.
  - Name match: The inventor's full name exactly matches the full name of a taxpayer on any type of information return form.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's information return forms.
  - 6% of patents are uniquely matched in this stage.

TABLE I
Summary Statistics

Sample:		Full	Intergenerational		New York City School District	
		Inventors	Inventors	Non-inventors	Inventors	Non-inventors
		(1)	(2)	(3)	(4)	(5)
Patenting Outcomes						
Patent Grants	Mean	3.0	1.4		1.3	
	Median	1.0	1.0		1.0	
	Std. Dev.	6.5	2.7		2.0	
Patent Applications	Mean	3.2	2.2		2.1	
	Median	1.0	1.0		1.0	
	Std. Dev.	11.1	4.3		3.4	
Patent Citations	Mean	26.2	1.2		1.3	
	Median	1.0	0.0		0.0	
	Std. Dev.	118.1	12.3		8.7	
Number of Collaborators	Mean	4.7	4.0		3.5	
	Median	2.0	3.0		2.0	
	Std. Dev.	8.2	5.0		4.4	
Age at Application	Mean	43.7	27.5		27.7	
	Std. Dev.	11.5	2.3		2.7	
Income in 2012						
Individual Wage Earnings ($)	Mean	111,457	82,902		94,622	
	Median	83,000	72,000		74,000	
	Std. Dev.	140,463	91,909		127,712	
Total Individual Income ($)	Mean	188,782	111,118		173,126	
	Median	100,000	74,000		75,000	
	Std. Dev.	567,813	396,673		800,082	
Parent Household Income ($)	Mean		183,303	85,992	108,049	47,509
	Median		109,000	59,000	66,000	33,000
	Std. Dev.		662,669	336,387	208,251	81,607
Attended College at Age 20			86.0%	47.7%		
Test Scores						
3rd Grade Mean Math Score					1.0	0.1
3rd Grade Mean English Score					0.8	0.1
8th Grade Mean Math Score					1.3	0.2
8th Grade Mean English Score					1.0	0.2
Demographics						
Female Share		13.1%	18.5%	49.8%	21.9%	48.8%
White Non-Hispanic Share					44.9%	19.5%
Black Non-Hispanic Share					17.3%	36.0%
Hispanic Share					8.4%	33.7%
Asian Share					27.4%	9.6%
Sample Size		1,200,689	34,973	16,360,910	452	433,863

Notes: This table presents summary statistics for the three samples of inventors and corresponding samples of non-inventors used in the empirical analysis. We define individuals as inventors if they were listed as an inventor on a patent application between 2001-2012 or grant between 1996-2014. The full inventors sample (Column 1) includes all inventors who were linked to the tax data using the procedure described in the Online Appendix. The intergenerational sample consists of U.S. citizens born in 1980-1984 matched to their parents in the tax data (Columns 2 and 3). The New York City School District sample includes children in the 1979-1985 birth cohorts who attended New York City public schools at some point between grades 3-8 and were linked to the tax data (Columns 4 and 5). Citations are measured as total patent citations between 1996-2014. The number of collaborators is measured as the number of distinct individuals that the inventor has ever co-authored a patent grant or application with in our linked dataset. For individuals with more than one patent application, age at application is the age at a randomly selected patent application filing. Incomes are measured in 2012. Individual wage earnings is defined as total earnings reported on an individual's W-2 forms. Total individual income is defined for tax filers as Adjusted Gross Income (as reported on the 1040 tax return) minus the spouse's W-2 wage earnings (for married filers). For non-filers, total individual income is defined as wage earnings. In this table only, wage earnings are top-coded at $1 million and total individual income is top-coded at $10 million. Parent income is measured as mean household income (AGI) between 1996-2000. Median income variables are rounded to the nearest thousand dollars. College attendance at age 20 is measured using 1098-T forms filed by colleges, as in Chetty et al. (2017). Test scores, which are based on standardized tests administered at the district level, are normalized to have mean zero and standard deviation one by year and grade. See Section II for further details on sample and variable definitions.

TABLE II

Fraction of Gap in Innovation by Parental Income Explained by Differences in 3rd Grade Test Scores

	Patent Rates for Children with Parents Below 80th Percentile	Patent Rates for Children with Parents Above 80th Percentile	High vs. Low Income Innovation Gap
	(1)	(2)	(3)
Raw Estimates	0.52	1.93	1.41
	(0.05)	(0.20)	(0.21)
Reweighted to Match 3rd Grade Scores of High-Income Children	0.96	1.93	0.97
	(0.07)	(0.20)	(0.21)
Gap in Innovation Explained by 3rd Grade Test Scores:			31.2%

Notes: This table shows how much of the gap in patent rates by parental income can be explained by 3rd grade math test scores. The statistics in this table are based on the children in the New York City public schools sample, which consists of children in the 1979-1985 birth cohorts who attended New York City public schools and were linked to the tax data. We divide children into two groups: those with parents in the top quintile of the income distribution within the New York City sample ("high-income children") and all other children in the sample ("low-income children"). We define a child as an inventor if he or she is listed as an inventor on a patent application between 2001-2012 or grant between 1996-2014 (see Section II.B). The first row of the table lists the fraction of children who become inventors among low-income (Column 1) and high-income children (Column 2) along with the differences between these two values (Column 3). In the second row of the table, Column 1 shows the patent rate that low-income children would have if they had the same math test scores as the high-income children. We calculate this counterfactual rate by dividing the math test score distribution into ventiles (twenty bins) and then calculating the patent rate for low-income children weighting by the number of high-income children in each of the twenty bins. Column 2 repeats the patent rates for high-income children, and Column 3 shows the gap between the high-income patent rate and the counterfactual low-income patent rate in Column 1. This adjusted gap can be interpreted as the difference in patent rates that would remain if test scores were identical across low- and high-income children. The percentage of the raw gap in innovation explained by 3rd grade test score is the percentage reduction in the gap from the raw to the reweighted estimates. Standard errors are reported in parentheses.

TABLE III

Exposure to Innovation from Parents' Colleagues: Children's Innovation Rates vs. Patent Rates in Father's Industry

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Fraction Inventing	Fraction Inventing in Patent Category	Fraction Inventing in Patent Sub-Category	Fraction Inventing in Patent Class	Fraction Inventing in Patent Class
Patent Rate in Father's Industry	0.250 (0.028)				
Patent Rate in Father's Industry in Same Category		0.163 (0.018)			
Patent Rate in Father's Industry in Same Sub-Category			0.155 (0.017)		
Patent Rate in Father's Industry in Same Class				0.078 (0.013)	0.0598 (0.0125)
Patent Rate in Father's Industry in Same Sub-Category but Other Class					0.0044 (0.0008)
Patent Rate in Father's Industry in Same Category but Other Sub-Category					0.0001 (0.0004)
Patent Rate in Father's Industry in Other Category					0.0002 (0.0000)
Fixed Effects	None	Patent Category	Patent Sub-Category	Patent Class	Patent Class
Unit of Observation	Father's Industry	Father's Industry by Patent Category	Father's Industry by Patent Sub-Category	Father's Industry by Patent Class	Father's Industry by Patent Class
Number of Cells	345	2,415	12,765	153,525	153,525
Mean of Dependent Variable	0.002341	0.000334	0.000063	0.000005	0.000005
SD of Dependent Variable	0.001063	0.000275	0.000118	0.000018	0.000018
Mean of Independent Variable	0.001040	0.000168	0.000034	0.000003	0.000003
SD of Independent Variable	0.002368	0.000654	0.000206	0.000030	0.000030

Notes: This table analyzes how a child's propensity to invent is related to patent rates in his or her father's industry. The sample consists of children in the intergenerational sample (1980-84 birth cohorts) whose parents are not inventors. Each column presents estimates from a separate OLS regression, with standard errors clustered by industry in parentheses. In Column 1, we regress the share of children who become inventors among those with fathers in industry j on the patent rate among workers in industry j, with one observation per industry (six digit NAICS code). We measure the patent rate among workers in each industry as the average number of patents issued to individuals in that industry per year between 1996-2012 divided by the average number of workers per year (based on W-2 counts) in each industry between 1999-2012. Column 2 is run at the industry by patent category level. Here, we regress the share of children with fathers in industry j who invent in patent category c on the share of workers in industry j who have patents in category c. We include patent category fixed effects in this regression to account for differences in patent rates across categories. Columns 3 and 4 are analogous to Column 2, but use more narrowly defined categorizations of patent types: patent sub-categories and patent classes. Column 5 replicates Column 4 with three additional controls: the fraction of inventors in (i) the same sub-category but in a different patent class, (ii) the same category but a different sub-category, and (iii) other categories. All regressions are weighted by the number of children in each cell. There are 10,213,731 children underlying these regressions, the set of children in the intergenerational sample whose fathers have a non-missing NAICS code.

TABLE IV
Neighborhood Exposure Effects: Children's Innovation Rates vs. Patent Rates in Childhood Commuting Zone

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	Fraction Inventing	Fraction Inventing	Fraction Inventing in Patent Category	Fraction Inventing in Patent Category	Fraction Inventing in Patent Category	Fraction Inventing in Patent Sub-Cat.	Fraction Inventing in Patent Class	Fraction Inventing in Patent Class
Patent Rate in Childhood CZ	2.932 (0.417)	2.578 (0.531)						
Patent Rate in Same Category in Childhood CZ			1.759 (0.404)	1.114 (0.341)	1.722 (0.406)			
Patent Rate in Same Sub-Category in Childhood CZ						1.526 (0.375)		
Patent Rate in Same Technology Class in Childhood CZ							1.108 (0.181)	1.017 (0.162)
Patent Rate in Same Sub-Category, but Different Technology Class in Childhood CZ								0.0003 (0.0063)
Patent Rate in Same Category, but Different Sub-Category in Childhood CZ								-0.0015 (0.0028)
Patent Rate in Different Category of Childhood CZ								0.0054 (0.0006)
Fixed Effects	None	Current CZ	Category	Current CZ by Category	Father's NAICS by Category	Sub-Category	Class	Class
Unit of observation	Childhood CZ	Childhood CZ by Current CZ	Childhood CZ by Category	Childhood CZ by Current CZ by Category	Childhood CZ by Father's NAICS by Category	Childhood CZ by Sub-Category	Childhood CZ by Patent Class	Childhood CZ by Patent Class
Number of Cells	741	221,621	5,187	1,551,347	1,637,706	27,417	329,745	329,745
Mean of Dep. Var.	0.002019	0.003692	0.000289	0.000527	0.000336	0.000055	0.000005	0.000005
SD of Dep. Var.	0.000905	0.010896	0.000240	0.003908	0.002477	0.000102	0.000017	0.000017
Mean of Indep. Var.	0.000286	0.000273	0.000041	0.000039	0.000042	0.000008	0.000001	
SD of Indep. Var.	0.000196	0.000204	0.000046	0.000048	0.000046	0.000013	0.000002	

Notes: This table analyzes how a child's propensity to invent is related to patent rates in his or her childhood commuting zone. The sample consists of children in the intergenerational sample (1980-84 birth cohorts) whose parents are not inventors. Each child is assigned a childhood CZ based on the ZIP code from which their parents first claimed them as dependents. Each column presents estimates from a separate OLS regression, with standard errors clustered by CZ in parentheses. In Column 1, we regress the share of children who become inventors among those who grow up in CZ j on the patent rate among workers in CZ j, with one observation per CZ. We measure the patent rate among workers in each CZ as the average number of patents issued per year (in the full USPTO data) to individuals in a given CZ between 1980 and 1990 divided by the CZ's population between the ages of 15-64 in the 1990 Census. Column 2 is run at the childhood CZ by current CZ level, limiting the sample to children whose current (2012) CZ differs from their childhood CZ. Here, we regress the share of inventors in each cell on the patent rate in the childhood CZ and on fixed effects for the 2012 CZ, so that the coefficient on childhood CZ patent rates is identified from comparisons across individuals currently living in the same CZ. Column 3 is run at the childhood CZ by patent category level. Here, we regress the share of children from CZ j who invent in patent category c on the share of workers in CZ j who have patents in category c. We include patent category fixed effects in this regression to account for differences in patent rates across categories. Column 4 replicates Column 2 at the category level, limiting the sample to children who move and estimating the model at the childhood CZ by current CZ by category level, with current CZ by category fixed effects. In Column 5, we include all children and replace the CZ by category fixed effects with fixed effects for the father's industry by category, estimating the model at the childhood CZ by father's industry by category level. This specification isolates variation from one's neighbors that is orthogonal to the variation from parents' colleagues. Columns 6 and 7 are analogous to Column 3 but use more narrowly defined categorizations of patent types: patent sub-categories and patent classes. Column 8 replicates Column 7 with three additional controls: the fraction of inventors in (i) the same sub-category but in a different patent class, (ii) the same category but a different sub-category, and (iii) other categories. All regressions are weighted by the number of children in each cell. There are approximately 15.5 million children underlying the regressions in Columns 1, 3, 6, 7 and 8. Columns 2 and 4 are based on the subset of 5.4 million individuals who moved across CZs. Column 5 includes the 10.2 million children whose fathers have non-missing NAICS codes.

TABLE V

Gender-Specific Exposure Effects: Children's Innovation Rates vs. Innovation Rates by Gender in Childhood CZ

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Fraction Inventing	Fraction of Women Inventing	Fraction of Men Inventing	Fraction of Women Inventing in Patent Category	Fraction of Men Inventing in Patent Category
Innovation Rate in Childhood CZ	0.986 (0.145)				
Innovation Rate of Women in Childhood CZ		2.408 (1.265)	-0.356 (4.398)	2.232 (0.607)	-2.157 (1.300)
Innovation Rate of Men in Childhood CZ		0.174 (0.154)	1.784 (0.625)	0.102 (0.062)	1.693 (0.295)
Fixed Effects	None	None	None	Category	Category
Unit of Observation	Childhood CZ	Childhood CZ	Childhood CZ	Childhood CZ by Category	Childhood CZ by Category
Number of Cells	741	741	741	5,188	5,188
p-value from F-test for Equality of Coefficients		0.113	0.667	0.001	0.015
Mean of Dep. Var.	0.002020	0.000745	0.003282	0.000102	0.000453
SD of Dep. Var.	0.000905	0.000396	0.001487	0.000117	0.000433
Mean of Indep. Var.	0.000628				
SD of Indep. Var.	0.000631				
Mean of Innov. Rate (Women)		0.000061	0.000060	0.000008	0.000008
SD of Innov. Rate (Women)		0.000066	0.000066	0.000017	0.000017
Mean of Innov. Rate (Men)		0.000568	0.000567	0.000080	0.000080
SD of Innov. Rate (Men)		0.000569	0.000568	0.000139	0.000139

Notes: This table analyzes how a child's propensity to invent is related to the innovation rates of adults of the same gender in his or her childhood commuting zone (CZ). The sample consists of children in the intergenerational sample (1980-84 birth cohorts) whose parents are not inventors. Each column presents estimates from a separate OLS regression, with standard errors clustered by CZ in parentheses. Column 1 replicates the specification in Column 1 of Table IV, except that here we define the independent variable using the linked patent-tax data rather than just the patent data, since we do not observe gender in the patent data itself. Specifically, we define the innovation rate for workers in CZ j as the total number of patent applications filed by individuals born before 1980 in our full inventors sample divided by the number of individuals between ages 15 and 64 in CZ j in the 1990 Census. We convert this measure to an annual rate by dividing by 17, as we observe patents between 1996-2012. In Column 2, we regress the fraction of girls from CZ j who become inventors on the patent rates of female and male workers in CZ j. Column 3 replicates Column 2 using the share of boys who become inventors as the dependent variable. The regression in column 4 is run at the childhood CZ by patent category level. Here, we regress the share of girls from CZ j who invent in patent category c on the share of male and female workers in CZ j who have patents in category c. We include patent category fixed effects in this regression to account for differences in patent rates across categories. Columns 5 replicates Column 4 using the share of boys who become inventors as the dependent variable. All regressions are weighted by the number of children in each cell. The last row of the table reports p-values from F-tests for equality of the coefficients on male and female innovation rates in each regression. There are 15,499,290 individuals underlying each of the regressions.

TABLE VI
Exposure Effects on Level of Innovation: Estimates Based on Movers

	(1)	(2)	(3)	(4)
Dependent Variable:	Indicator for Inventing by 2014			
Difference in Patent Rates x Age at Move	-0.0797 (0.0170)	-0.0806 (0.0164)	-0.0872 (0.0172)	
Difference in Patent Rates	3.137 (0.339)	3.166 (0.301)	4.612 (0.631)	
Origin Patent Rate		3.123 (0.146)	6.369 (0.671)	2.044 (0.282)
Origin FE	x			
Cohort FE x Difference in Patent Rates			x	
Cohort FE x Origin Patent Rate			x	
Age at Move FE	x	x	x	
Cohort FE	x	x	x	
N	3,637,481	3,637,481	3,637,481	28,798,471
Mean of Dep. Var.	0.00139	0.00139	0.00139	0.00138
SD of Dep. Var.	0.03732	0.03732	0.03732	0.03707
Mean of Difference in Patent Rates	-0.00004	-0.00004	-0.00004	
SD of Difference in Patent Rates	0.00026	0.00026	0.00026	
Mean of Origin Patent Rate		0.00028	0.00028	0.00029
SD of Origin Patent Rate		0.00020	0.00020	0.00020

Notes: This table analyzes how a child's propensity to innovate varies with the amount of time spent during childhood (before age 24) in a neighborhood with a low vs. high fraction of inventors among adults in the area. The sample consists of children in an extended intergenerational sample (1980-88 birth cohorts) whose parents are not inventors. Each column presents estimates from a separate OLS regression run at the individual level. The dependent variable in each regression is an indicator for whether the child is an inventor. Columns 1-3 include children whose parents moved across CZs exactly once between 1996 and 2014. Children's origin and destination CZ's are coded based on the ZIP codes from which their parents filed taxes in each year. As in Table IV, each CZ's mean patent rate among adults is defined as the average number of patents issued per year (in the full USPTO data) to individuals in a given CZ between 1980 and 1990 divided by the CZ's population between the ages of 15-64 in the 1990 Census. The variable "Difference in Patent Rates" is the patenting rate of adults in the destination CZ minus that in the origin CZ. Age at move refers to the child's age at time of the parent's move; if the age at move is above 24, it is recoded to 24 given Chetty and Hendren's (2018) finding that neighborhood exposure matters only up to age 24. The youngest moves in this sample occur at age 9 and the oldest, prior to recoding, at 32. The coefficient on "Difference in Patent Rates x Age at Move" can be interpreted as the causal effect of one additional year of exposure to a higher-innovation area (i.e., an area with higher observed patent rates) during childhood. Column 1 includes indicators for the child's birth cohort and age at move as well as origin CZ fixed effects as additional controls. Column 2 controls for origin patent rates among adults rather than origin fixed effects. Column 3 shows robustness of the estimates to interacting the controls in Column 2 with birth cohort. Finally, Column 4 replicates the specification in Column 1 of Table IV in the extended intergenerational sample as a reference. Here, we regress an indicator for being an inventor on the patent rates of adults in the first CZ in which we observe the child, which we call the Origin CZ for the purpose of this table. Standard errors, reported in parentheses, are unclustered in Columns 1-3 and are clustered by Origin CZ in Column 4.

APPENDIX TABLE I

Association Between Patent Rates and Upper-Tail Incomes with 3rd grade Math vs. English Test Scores

Dependent Variable:	Inventor (per 1,000 Individuals)					In Top 1% of Income Distribn. (per 1,000 Individuals)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3rd Grade Math Score (SD)	0.85***		0.85***	0.91***		11.07***		8.08***
	(0.09)		(0.13)	(0.13)		(0.36)		(0.46)
3rd Grade English Score (SD)		0.68***	0.05		0.08		10.19***	4.15***
		(0.09)	(0.11)		(0.12)		(0.34)	(0.42)
Fixed Effects	None	None	None	English Ventile	Math Ventile	None	None	None
Mean of Dependent Variable	0.83	0.86	0.86	0.86	0.86	9.68	9.72	9.84
Observations	222,369	214,265	209,722	209,722	209,722	165,422	161,275	158,016

Notes: This table examines the extent to which third grade test scores are predictive of patent rates and upper-tail earnings outcomes. The sample in columns 1-5 consists of children in the 1979-1985 birth cohorts who attended New York City public schools in 3rd grade and were linked to the tax data. The sample in columns 6-8 consists of children who appear in both the NYC school district and intergenerational samples (1980-84 birth cohorts). Each column shows the coefficients and robust standard errors (in parentheses) from a separate OLS regression run at the student level; *** denotes $p < 0.001$. In Columns 1-5, the dependent variable is an indicator for being an inventor, defined as applying for a patent between 2001-2012 or being granted a patent between 1996-2014. In columns 6-8, it is an indicator for being in the top 1% of the individual income distribution in 2012 when compared to other individuals in the same birth cohort in the NYC school district sample. The dependent variables in each column are math and English test scores in 3rd grade. Test scores, which are based on standardized tests administered at the district level, are normalized to have mean zero and standard deviation one by year and grade. In Columns 4 and 5, we control for English and math scores non-parametrically using ventile fixed effects (20 bins) rather than a linear control. In all columns, coefficients are scaled so that they can be interpreted as the effect of a 1 SD change in test scores on the number of individuals per 1,000 who have the relevant outcome.

APPENDIX TABLE II

Fraction of Gender Gap in Innovation Explained by Differences in Test Scores

A. Percent of Innovation Gap Explained by 3rd Grade Math Test Scores

	Patent Rates for Women	Patent Rates for Men	Gender Gap
	(1)	(2)	(3)
Raw Estimates	0.43	1.13	0.70
	(0.06)	(0.09)	(0.11)
Reweighted to Match 3rd Grade Scores of Men	0.45	1.13	0.68
	(0.06)	(0.09)	(0.11)
	Gap in Innovation Explained by 3rd Grade Test Scores:		2.4%

B. Percent of Gap Explained by Test Scores Grades 3-8

Grade	Percent of Innovation Gap Explained by Math Test Scores in Grade *g*
3	2.4%
	(0.4)
4	2.4%
	(0.3)
5	3.4%
	(0.4)
6	4.6%
	(0.5)
7	6.8%
	(0.8)
8	8.5%
	(1.0)
Slope:	1.3
	(0.2)

Notes: This table shows how much of the gender gap in patent rates can be explained by test scores using the New York City school district sample. Panel A is constructed in exactly the same way as Table II, comparing girls with boys instead of low-income children with high-income children. Panel B presents estimates of the gender gap in innovation that can be explained by test scores in grades 3-8, analogous to the estimates in Figure V. The slope estimate reported at the bottom is estimated using an OLS regression of the six estimates on grade.

APPENDIX TABLE III

Distance Between Technology Classes: Illustrative Example

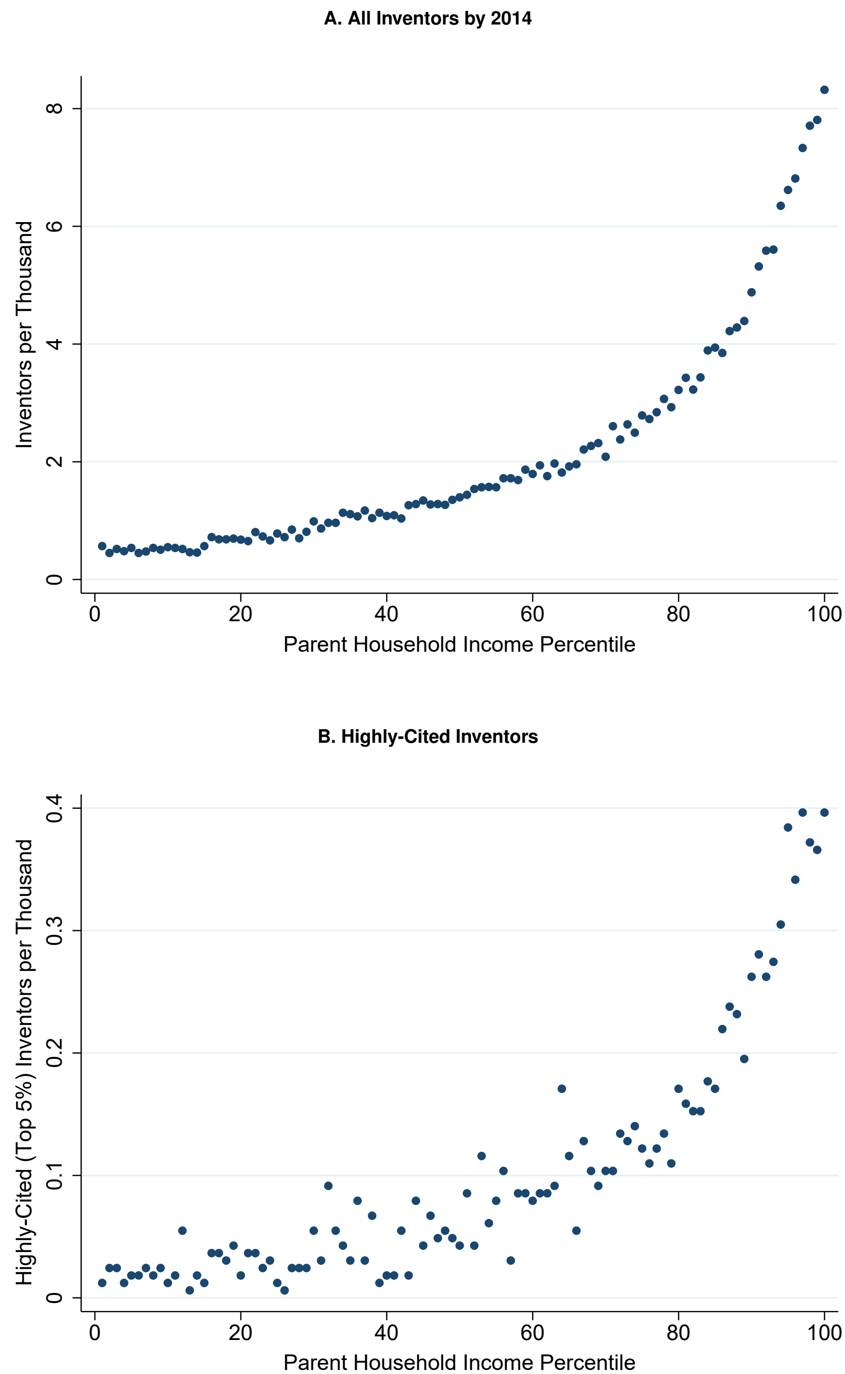
Category: Computers + Communications

Sub-category*: Communications*

Technology Class (= 375)	Distance Rank (*d*)
Pulse or Digital Communications	0
Demodulators	1
Modulators	2
Coded Data Generation or Conversion	3
Electrical Computers: Arithmetic Processing and Calculating	4
Oscillators	5
Multiplex Communications	6
Telecommunications	7
Amplifiers	8
Motion Video Signal Processing for Recording or Reproducing	9
Directive Radio Wave Systems and Devices (e.g., Radar, Radio Navigation)	10

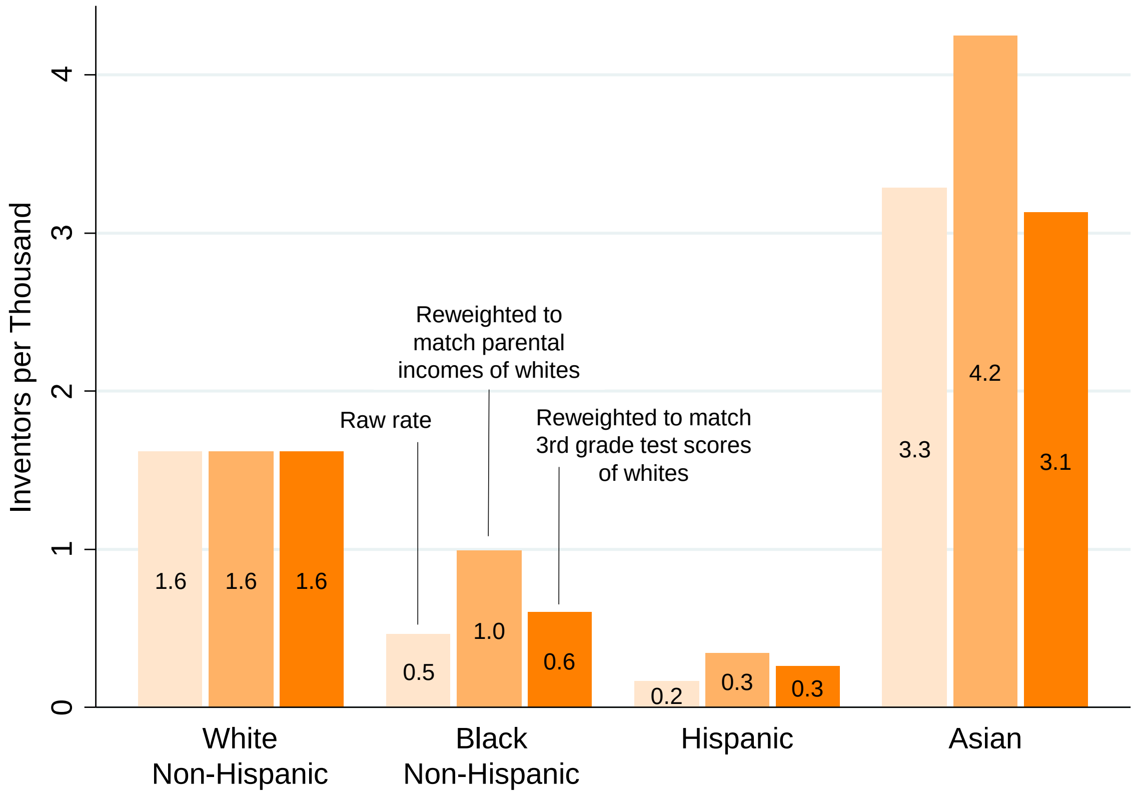
Notes: This table provides an example of our measures of distance between technology classes. We define the distance between two technology classes A and B by computing the share of inventors in class A who also invent in class B; the higher the share of common inventors, the lower the distance between A and B. We convert this distance metric to an ordinal measure, defining d=0 for the own class, d=1 for the next nearest class, etc. The table lists the 10 closest classes to the "Pulse or Digital Communtications" class, which falls within the Communications subcategory of the Computers + Communications category.

FIGURE I: Patent Rates vs. Parent Income



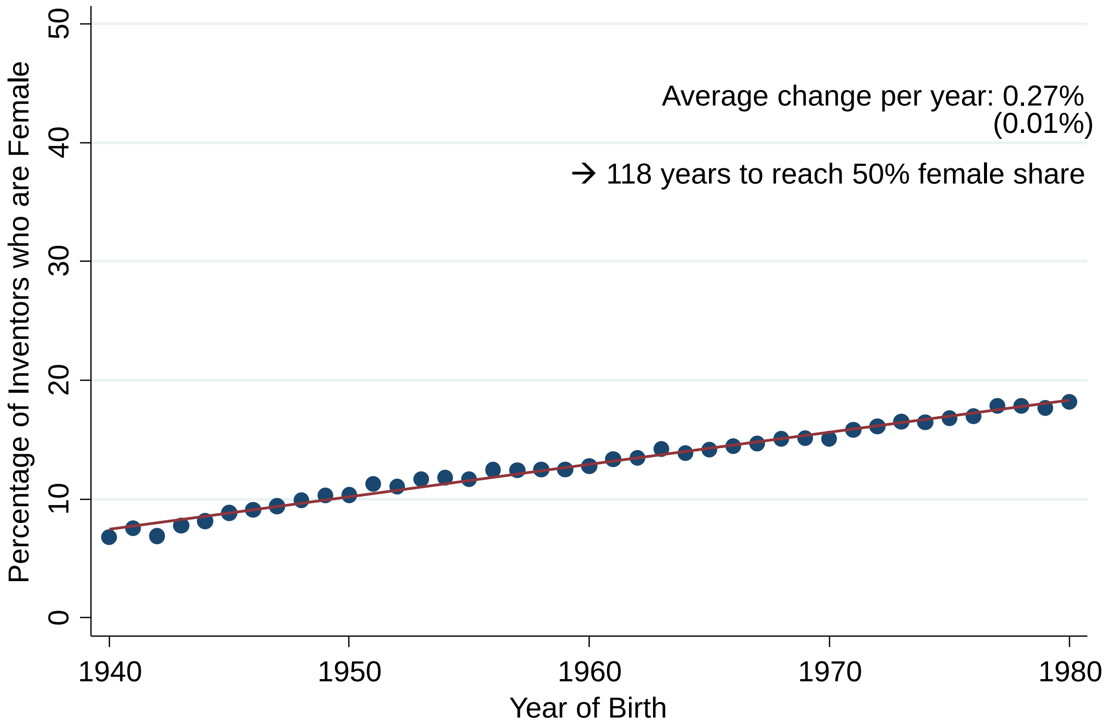
Notes: This figure characterizes the relationship between patent rates and parental income using our intergenerational analysis sample, which consists of U.S. citizens in the 1980-84 birth cohorts (see Section II.B for details). Panel A plots the number of children (per 1,000 individuals) who invent by 2014 vs. their parents' income percentile. Parents are assigned percentile ranks by ranking them based on their mean household income from 1996 to 2000 relative to other parents with children in the same birth cohort. Inventing by 2014 is defined as being listed as an inventor on a patent application between 2001-2012 or grant between 1996-2014 (see Section II.B). Panel B replicates Panel A, but plots as the outcome the chances of being a highly-cited inventor, defined as having total citations in the top 5% of the distribution among inventors in the same birth cohort.

FIGURE II: Patent Rates by Race and Ethnicity



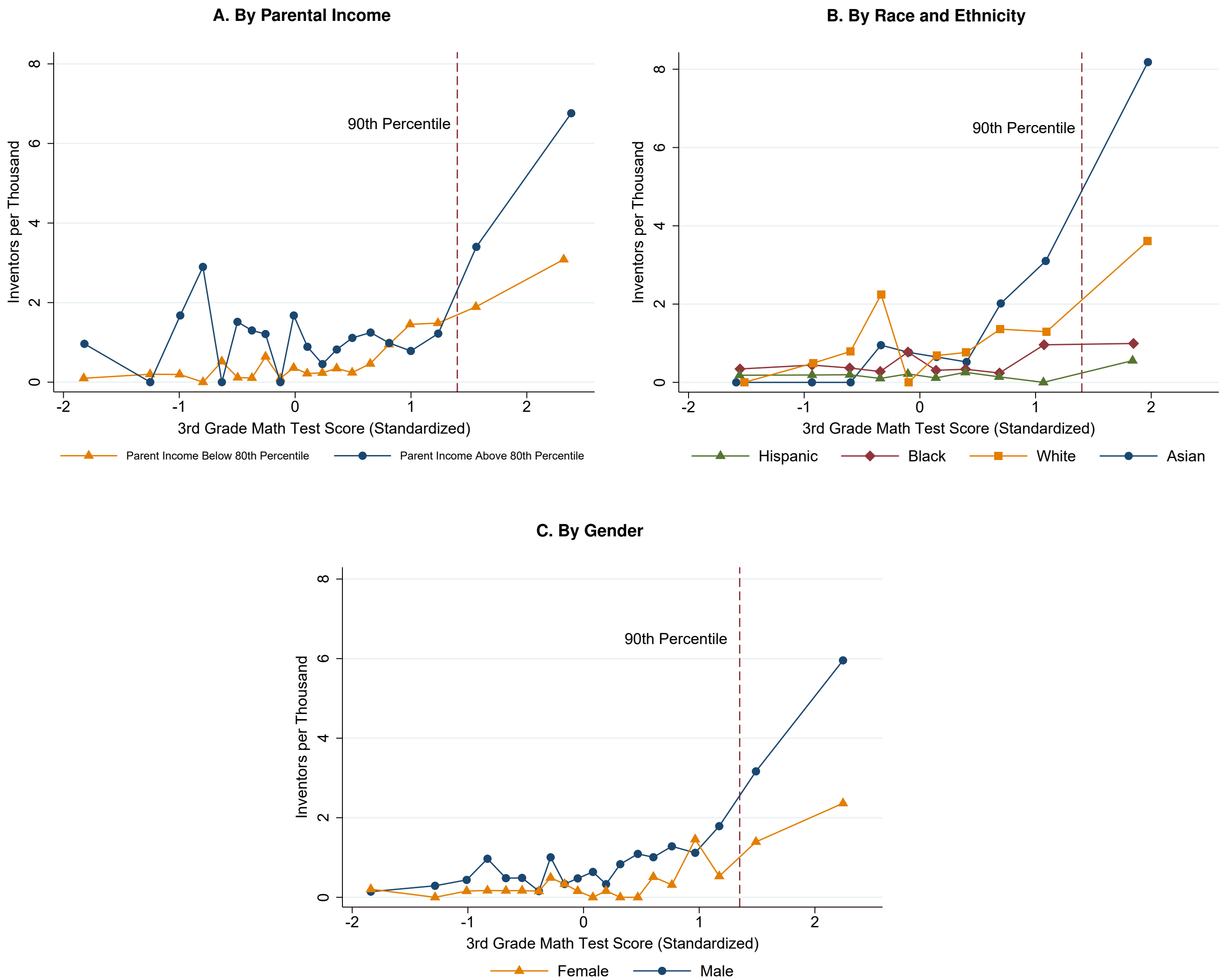
Notes: This figure presents patent rates by race and ethnicity using our New York City public schools sample, which consists of children in the 1979-1985 birth cohorts who attended NYC public schools at some point between grades 3-8. Each bar plots the number of children (per 1,000 individuals) who invent by 2014, as defined in the notes to Figure I. In each triplet, the first bar shows the raw patent rate for the relevant subgroup. The second bar plots the patent rate that would prevail if children in the relevant subgroup had the same distribution of parental income as white children. To construct these estimates, we divide children into 20 bins based on their parental incomes and compute mean patent rates across the 20 bins, weighting each bin by the fraction of white children with incomes in that bin. The third bar in each triplet shows the patent rate that would prevail if children in the relevant subgroup had the same distribution of 3rd grade math test scores as white children. These estimates are constructed by dividing children into 20 bins based on their test scores and computing mean patent rates across the 20 bins, weighting each bin by the fraction of white children with test scores in that bin.

FIGURE III: Percentage of Female Inventors by Birth Cohort



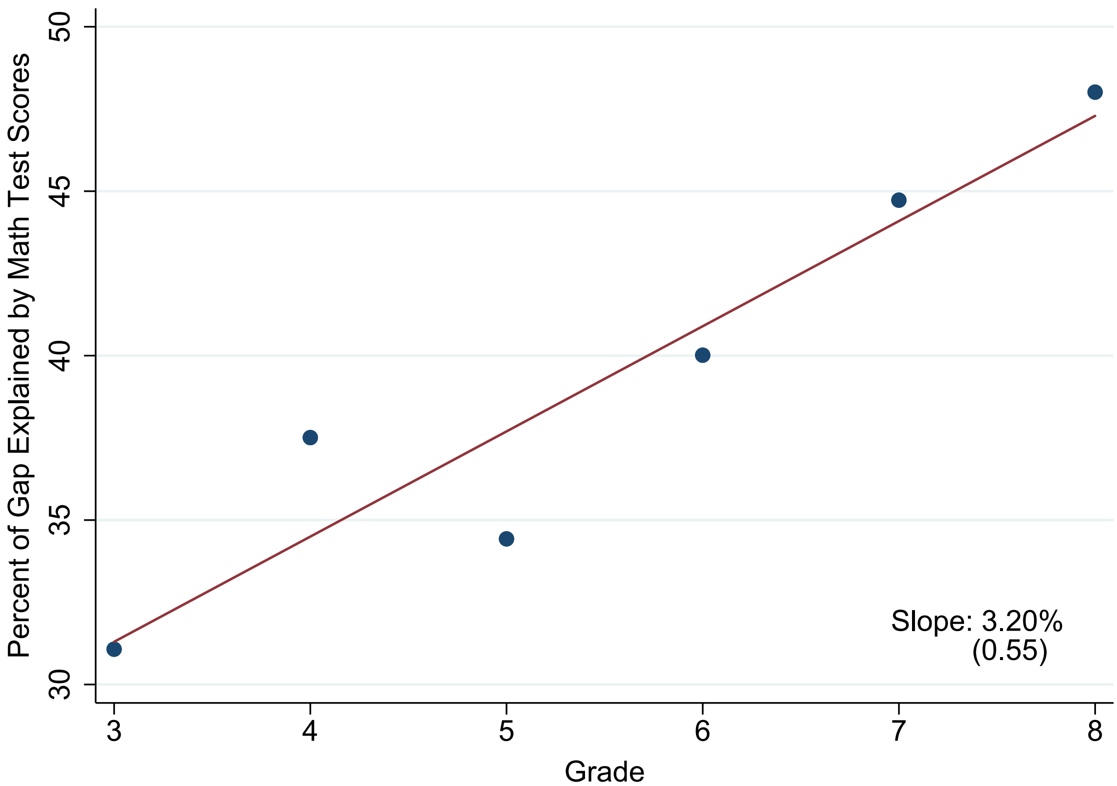
Notes: This figure plots the percentage of inventors who are female by year of birth using our full inventors sample, which consists of all 1.2 million individuals in the linked patent-tax data. Inventing is defined as being listed as an inventor on a patent application between 2001-2012 or grant between 1996-2014 (see Section II.B for details). The change per year is estimated using an unweighted OLS regression of the percentage of female inventors on birth year, depicted by the solid line. The standard error from this regression is shown in parentheses.

FIGURE IV: Patent Rates vs. 3rd Grade Math Test Scores



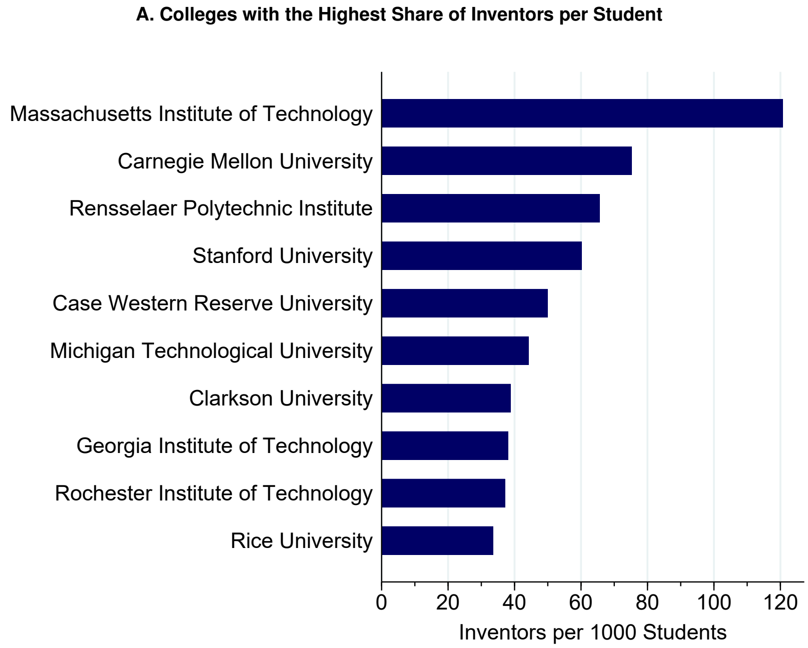
Notes: This figure shows the relationship between patent rates and math test scores in 3rd grade for various subgroups. The sample consists of children in the 1979-1985 birth cohorts who attended New York City public schools in 3rd grade. Test scores, which are based on standardized tests administered at the district level, are normalized to have mean zero and standard deviation one by year and grade. In Panel A, we divide children into two groups based on whether their parents' incomes fall below the 80th percentile of the income distribution of parents' income in the New York City schools sample. The figure presents a binned scatter plot of patent rates vs. test scores for these two subgroups. To construct the figure, we first divide children into 20 equal sized bins (ventiles) based on their test scores. We then plot the share of inventors (per 1,000 individuals) vs. the mean test score within each bin for each of the two subgroups. Panel B and C replicate Panel A, dividing children by their race and ethnicity (Panel B) and gender (Panel C) instead of parental income. We use 10 bins rather than 20 bins of test scores in Panel B because of smaller sample sizes for some racial and ethnic groups. The vertical dashed lines depict the 90th percentile of the test-score distribution.

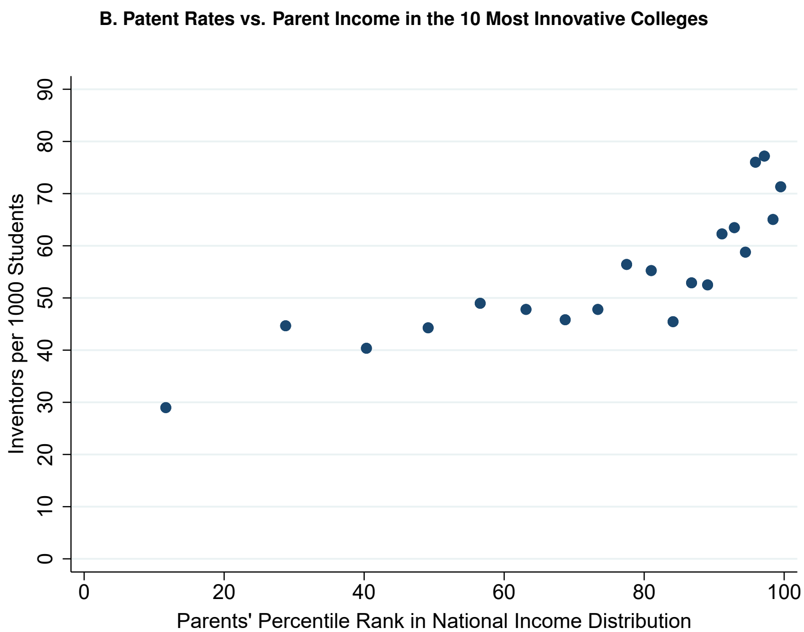
FIGURE V: Gap in Patent Rates by Parental Income Explained by Test Scores in Grades 3-8



Notes: This figure shows how much of the gap in patent rates by parental income can be explained by math test scores in grades 3-8. The sample consists of children in our New York City public schools sample (birth cohorts 1979-1985), who we divide into two groups: those with parents in the top quintile of the income distribution within the New York City sample ("high-income children") and all other children in the sample ("low-income children"). The gap in innovation explained by math test scores in grade g is the percentage reduction in the gap in innovation when we reweight low-income students' grade g test score distribution to match that of high-income students. Table II illustrates how we construct this estimate using 3rd grade test scores (31.2%); estimates for later grades use the same methodology. The slope and best-fit line are estimated using an unweighted OLS regression on the six points, with standard error reported in parentheses.

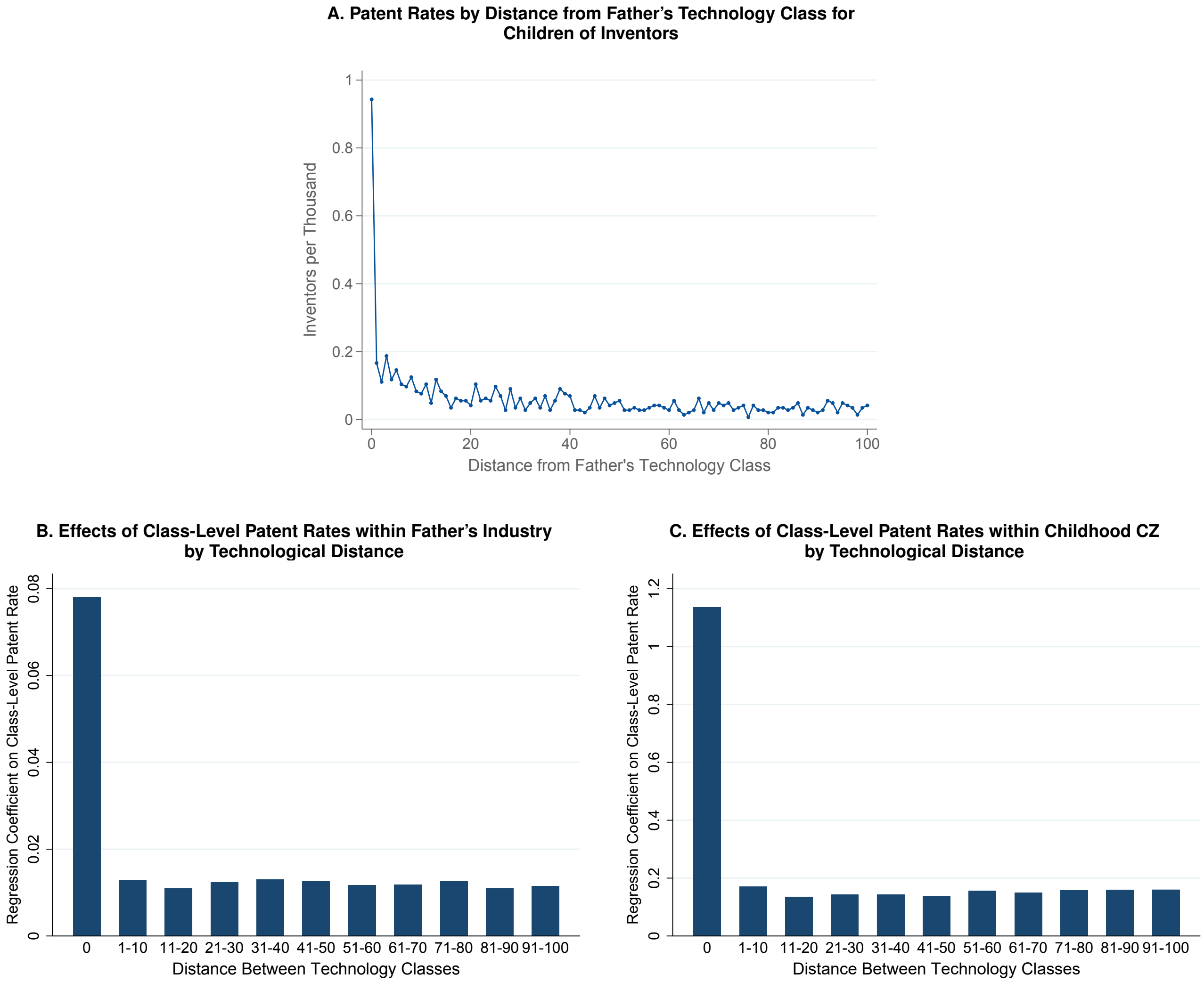
FIGURE VI: Patent Rates by College





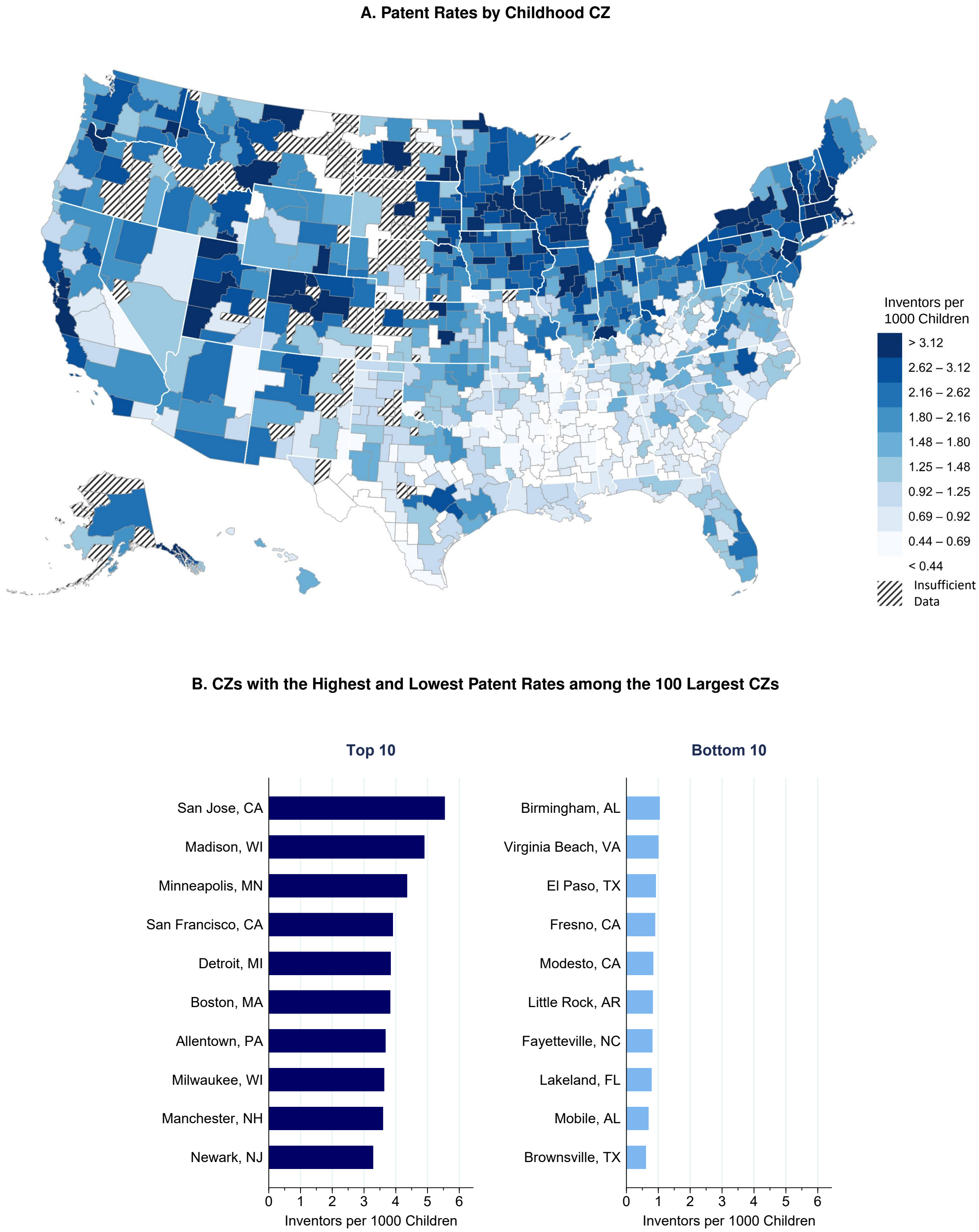
Notes: This figure presents data on the share of students who become inventors by 2014 (as defined in the notes to Figure I) by the college they attended. The sample consists of all individuals in the tax data in the 1980-84 birth cohorts who are linked to parents. Children are assigned to the college that they attend most frequently at age 19-22, following the methodology of Chetty et al. (2017). Panel A lists the ten colleges that have the highest fraction of students who become inventors, among colleges with at least 500 students per cohort. This figure is produced from the college-level estimates in Online Data Table 3. These college-level estimates are blurred to protect confidentiality using the procedure in Chetty et al. (2017, Appendix C). Panel B presents a binned scatterplot of patent rates vs. parental income for students who attended the 10 colleges listed in Panel A. It is constructed by binning parent income into 20 equal-sized bins (ventiles) and plotting the mean share of inventors (per 1,000 students) vs. the mean parent rank in the national income distribution within each bin. There are fewer points on the left because there are fewer students from low-income families than high-income families at these colleges.

FIGURE VII: Children's Patent Rates vs. Class-Level Patent Rates in Childhood Environment



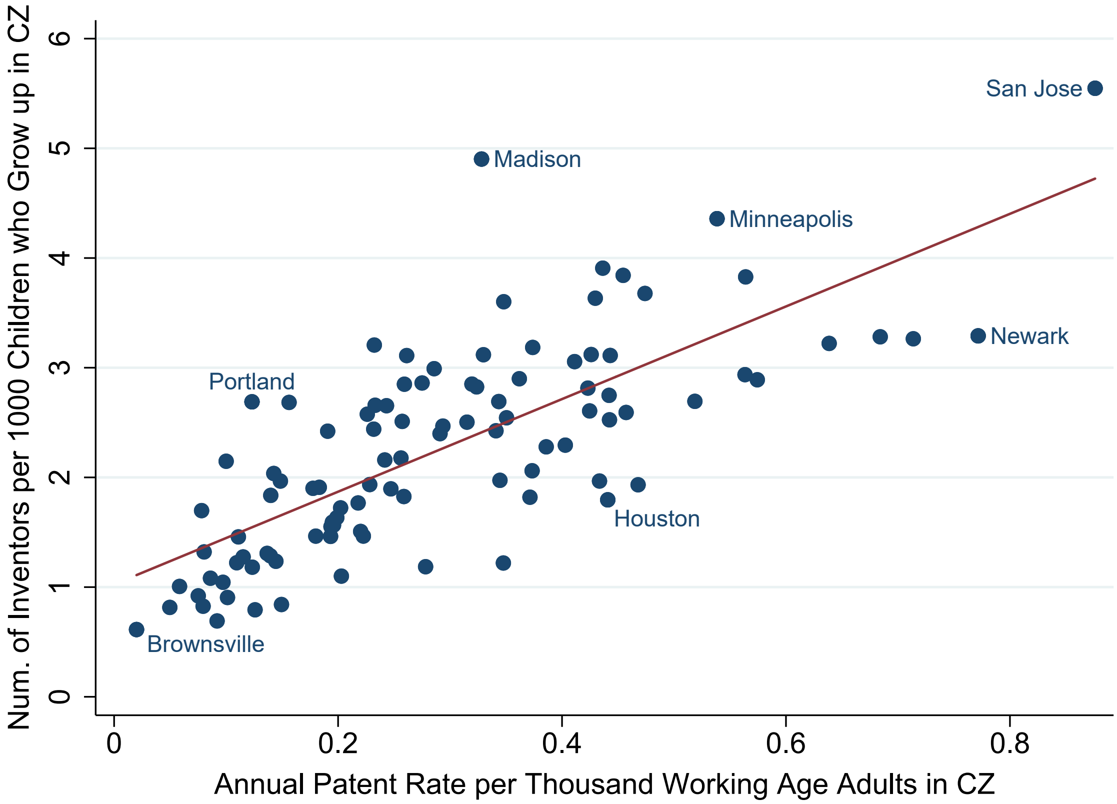
Notes: This figure shows how children's propensities to patent in a technology class vary with the class in which their father (Panel A), father's colleagues (Panel B), or childhood neighbors (Panel C) patented. In Panel A, the sample consists of all children in our intergenerational sample whose fathers are inventors (those who applied for a patent between 2001-2012 or were granted a patent between 1996-2014) and who were not listed as co-inventors on a patent with their fathers. To construct Panel A, we first assign fathers and children a technology class based on the class in which they have the most patents and patent applications. We then define the distance between two technology classes A and B based on the share of inventors in class A who also invent in class B. Using this distance metric, for each child, we define $d = 0$ as the class in which his or her father patents, $d = 1$ as the next closest class, etc. We then plot the share of children (per 1,000 individuals) who invent in a technology class that is d units away from their father's class. Classes in which fewer than 100 inventors have a patent grant or application between 1996-2014 are omitted. In Panels B and C, the sample consists of all children in our intergenerational sample whose parents are *not* inventors. Each bar in Panel B plots estimates from a separate regression, with one observation per father's industry (six digit NAICS code) and patent technology class. In the first bar, we regress the fraction of children who patent in technology class c among those with fathers in industry j on the patent rate among workers in industry j in the same technology class c. We measure the class-level patent rate among workers in each industry as the average number of patents in class c issued to individuals in that industry per year (between 1996-2012) divided by the average number of workers per year in each industry between 1999-2012. In the second bar, we regress the same dependent variable on the mean patent rate in the father's industry in the 10 closest classes ($d = 1$ to 10). The third bar uses the average patent rate in classes with $d = 11$ to 20, etc. All regressions are weighted by the number of children in each cell and include class level fixed effects for class c. Panel C replicates Panel B, replacing patent rates in the father's industry with patent rates of workers in the CZ where the child grew up. CZ-level patent rates are defined as the average number of patents issued in class c per year to individuals from a given CZ between 1980-1990 divided by the CZ's population between ages 15-64 in the 1990 Census.

FIGURE VIII: The Origins of Inventors: Patent Rates by Childhood Commuting Zone



Notes: Panel A maps the share of children who become inventors by the commuting zone (CZ) in which they grew up using our intergenerational sample (U.S. citizens in the 1980-84 birth cohorts). Each child is assigned a CZ based on the ZIP code from which their parents filed their 1040 tax return in the year they were first claimed as dependents (which is typically 1996, as our data begin in 1996). The map is constructed by dividing the CZs into unweighted deciles based on patent rates, with darker shades representing areas where more children grow up to become inventors. Data for CZs with fewer than 1,000 children, which account for 0.3% of the children in the sample, are omitted. Panel B lists the CZs with the ten highest and lowest shares of inventors per thousand children among the 100 CZs with the largest populations in the 2000 Census.

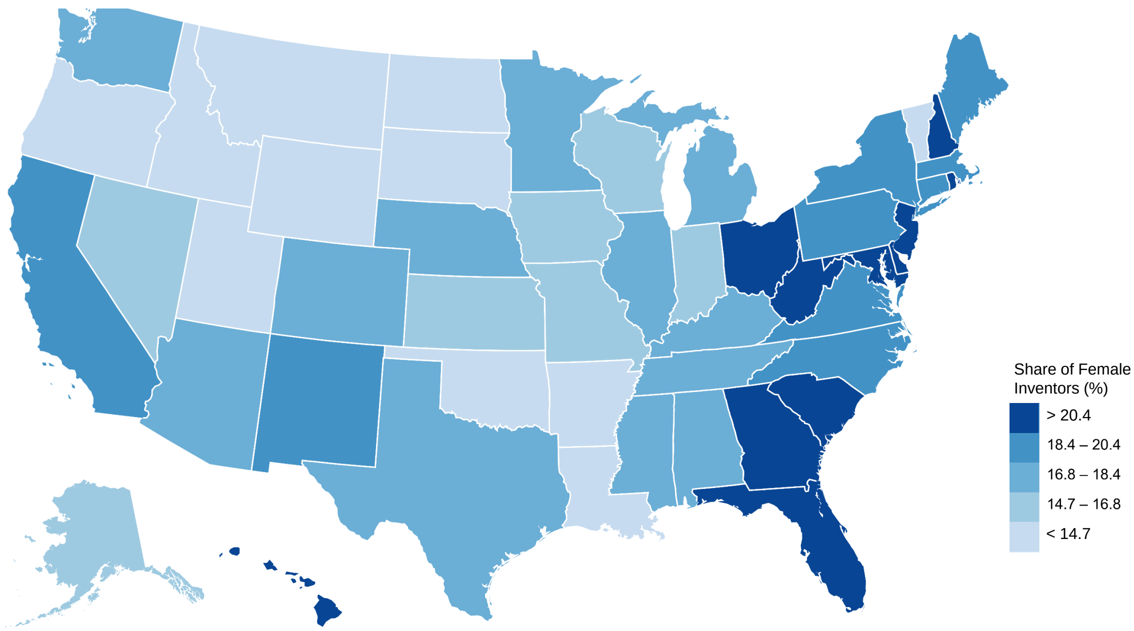
FIGURE IX: Children's Patent Rates vs. Patent Rates of Workers in their Childhood CZ



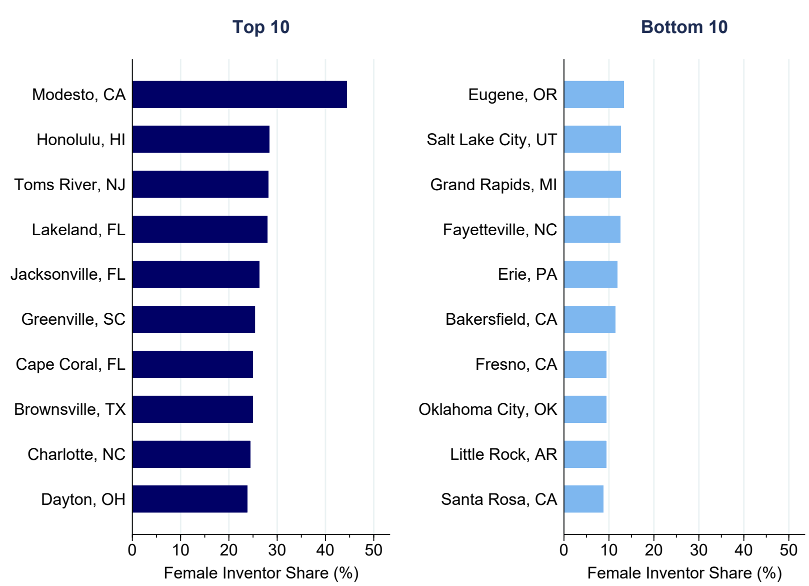
Notes: The figure plots the patent rates of children who grow up in a given CZ (constructed exactly as in Figure VIII) vs. the patent rates of workers who live in that CZ. Patent rates of workers in each CZ are defined as the average number of patents per year issued to inventors residing in that CZ between 1980-1990 (based on the universe of USPTO data) divided by the CZ's population between the ages of 15-64 in the 1990 Census. We restrict the figure to the 100 CZs with the largest populations in the 2000 Census. The solid best-fit line is estimated using an unweighted OLS regression on these 100 observations (slope = 4.22, standard error = 0.40).

FIGURE X: Geographical Variation in Gender Gaps in Patent Rates

A. Percent of Inventors who are Female by State where Child Grew Up

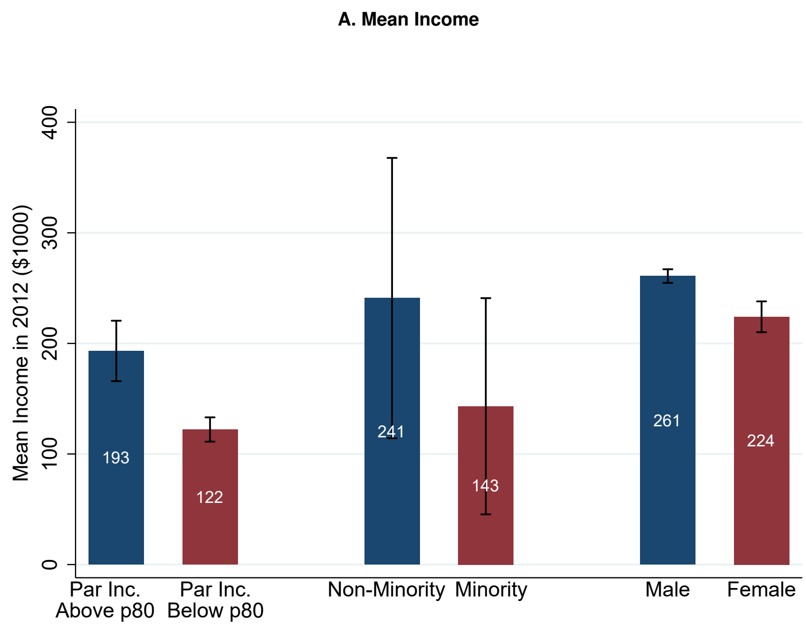


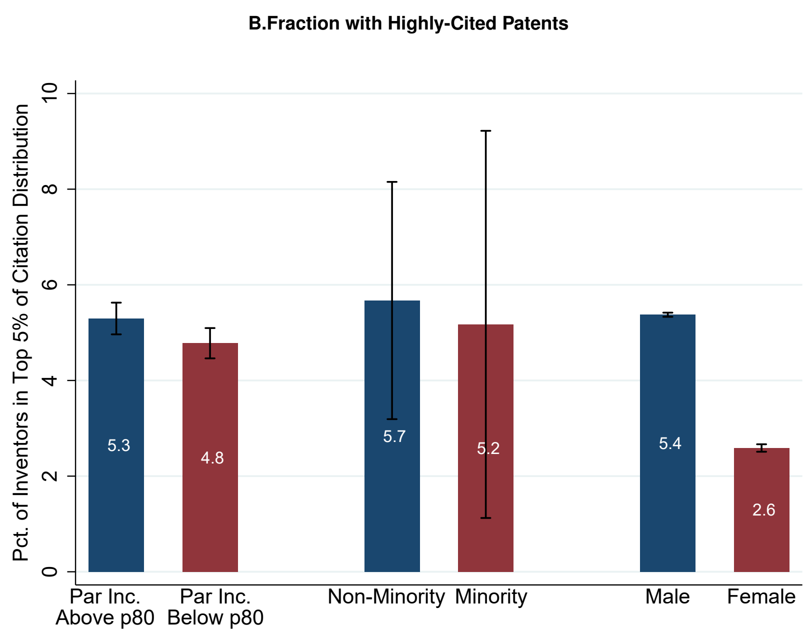
B. CZs with the Highest and Lowest Female Inventor Shares Among the 100 Largest CZs



Notes: Panel A maps the percentage of female inventors by the state in which they grew up using our intergenerational sample (U.S. citizens in the 1980-84 birth cohorts). Each child is assigned a state based on ZIP code from which their parents filed their 1040 tax return in the year they were first claimed as dependents (which is typically 1996, as our data begin in 1996). The map is constructed by dividing the states into unweighted quintiles based on the female inventor share, with darker shades representing areas where women account for a larger share of inventors. Panel B lists the commuting zones (CZs) with the ten highest and lowest female inventor shares among the 100 CZs with the largest populations in the 2000 Census.

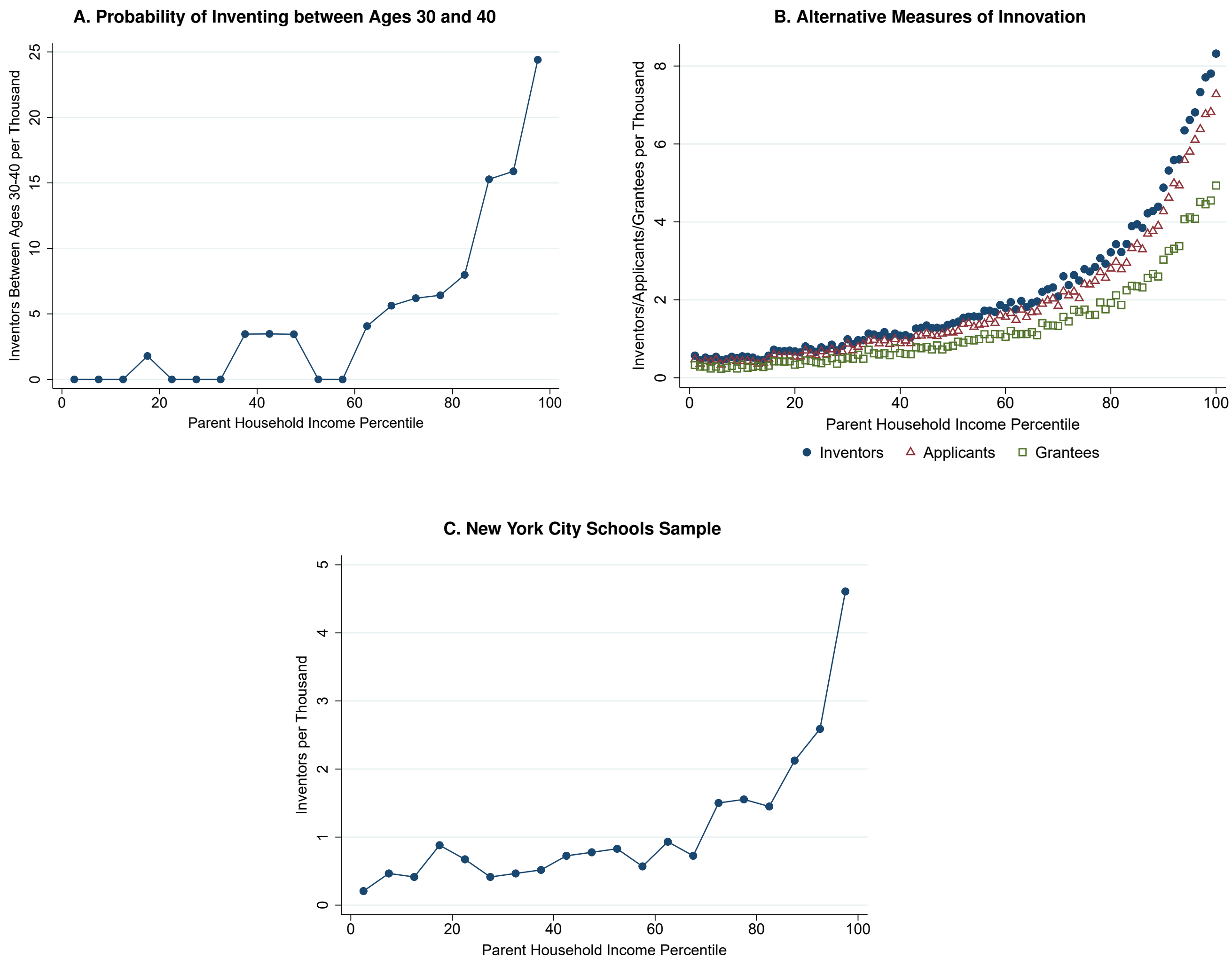
FIGURE XI: Income and Citations of Inventors by Characteristics at Birth





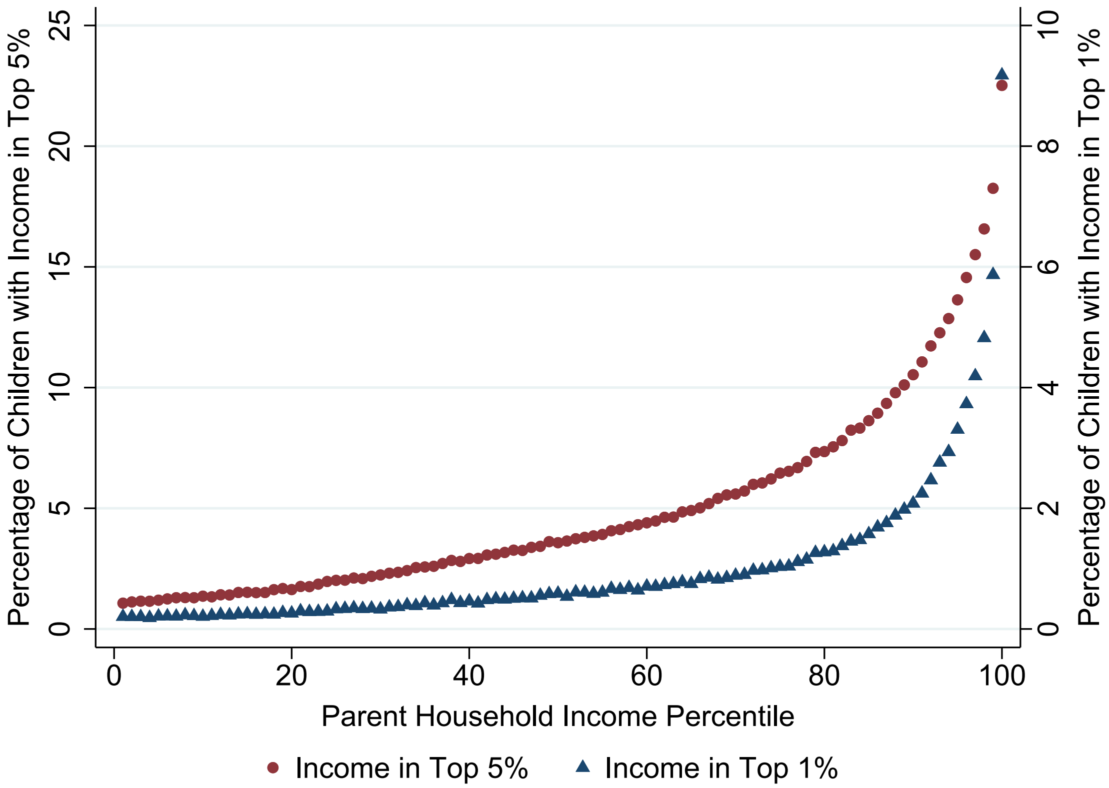
Notes: This figure presents how two measures of inventor productivity (income and citations) differ across various demographic groups. Panel A plots the mean incomes of inventors in 2012 by their parents' income, race/ethnicity, and gender. The first pair of bars uses our intergenerational sample (1980-84 birth cohorts), divided into two subgroups based on whether parents' household income is below or above the 80th percentile of the parent income distribution. The second pair of bars uses our New York City schools sample, divided into two subgroups based on race and ethnicity: minorities (Blacks and Hispanics) and non-Minorities. The third pair of bars uses our full inventors sample, divided by gender. The vertical lines depict 95% confidence intervals. Panel B replicates Panel A using the fraction of highly-cited inventors as the outcome. Highly-cited inventors are defined as inventors whose patents have citations per co-author in the top 5% of the distribution among those in their birth cohort.

ONLINE APPENDIX FIGURE I: Patent Rates vs. Parent Income: Sensitivity Analysis



Notes: This figure replicates Figure Ia using alternative samples and definitions. Panel A uses data from the 1971-72 birth cohorts in the Statistics of Income sample, a 0.1% sample of tax returns (see Section II.B for details). This sample allows us to examine whether an individual filed a patent application or was granted a patent between the ages of 30 and 40, rather than just by their early thirties as in Figure Ia. Given the smaller sample size, we plot mean patent rates between the ages of 30-40 by parent ventile (20 bins) rather than percentiles in this figure. In Panel B, the series in circles replicates Figure Ia exactly, where inventors are defined as those who applied for a patent between 2001-2012 or were granted a patent between 1996-2014. The other two series in that figure show the fraction of individuals who applied for patents and the fraction who were granted patents separately. Panel C replicates the baseline series in Figure Ia (plotting the fraction of inventors) using the subset of children in the New York City public schools sample. In this figure, we rank parents within the NYC sample based on their household incomes and plot the fraction of children who become inventors by 2014 by parent income ventiles.

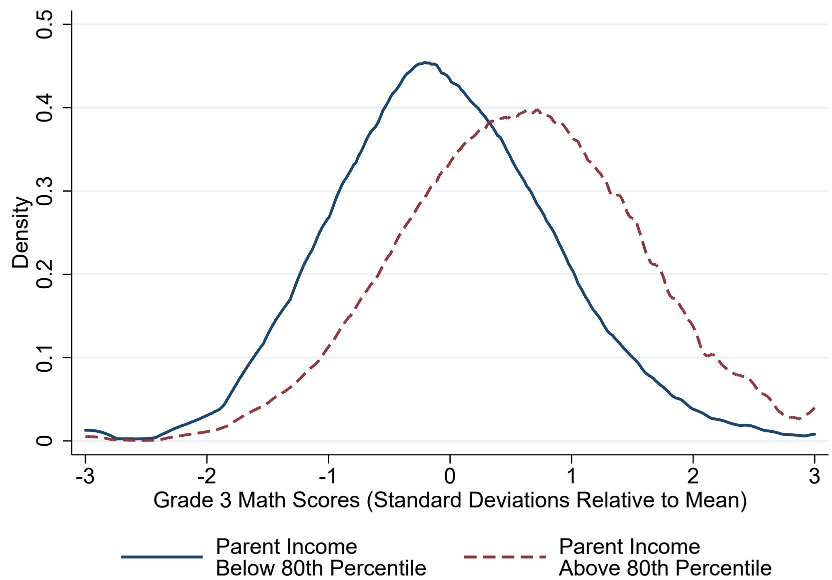
ONLINE APPENDIX FIGURE II:
Fraction of Children with Incomes in Upper Tail vs. Parent Income



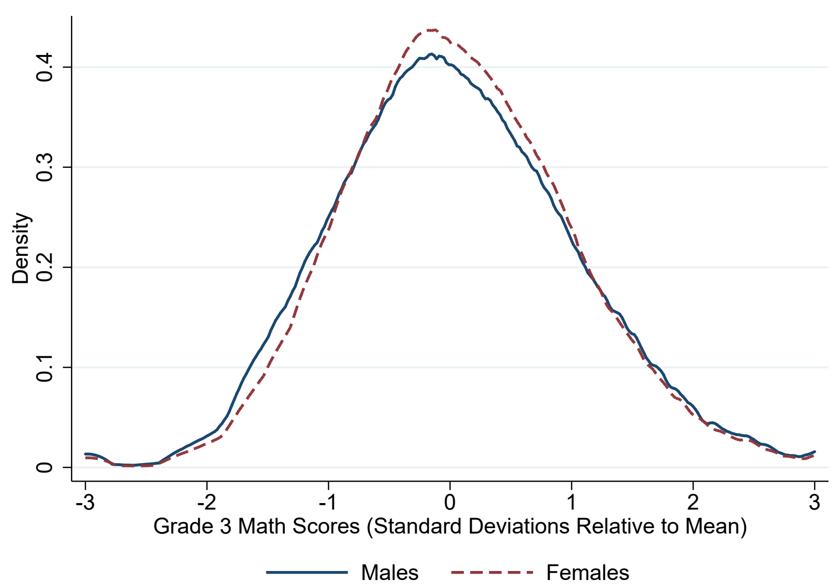
Notes: This figure replicates Figure I, replacing the outcome variable with an indicator for having mean individual income in 2011-12 in either the top 1% or top 5% of the income distribution among individuals in the same birth cohort. The sample is our core intergenerational sample of the 1980-84 birth cohorts.

ONLINE APPENDIX FIGURE III: Distribution of Math Test Scores in 3rd Grade

A. By Parental Income

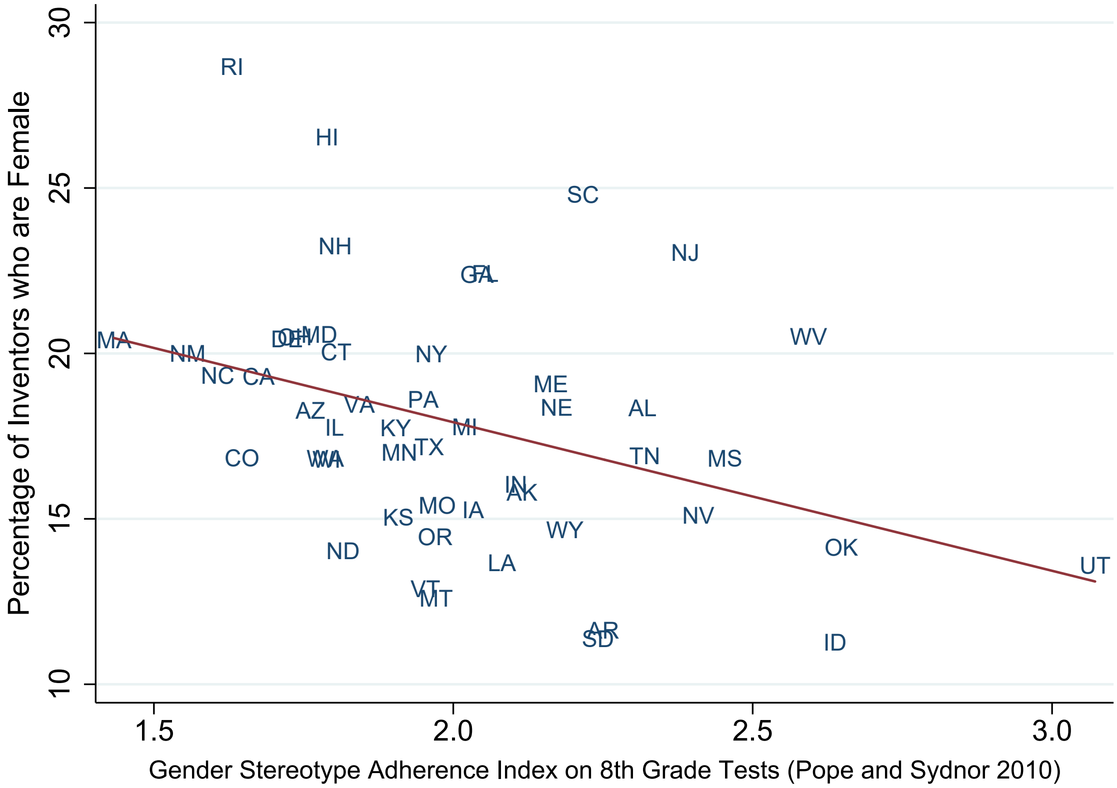


B. By Gender



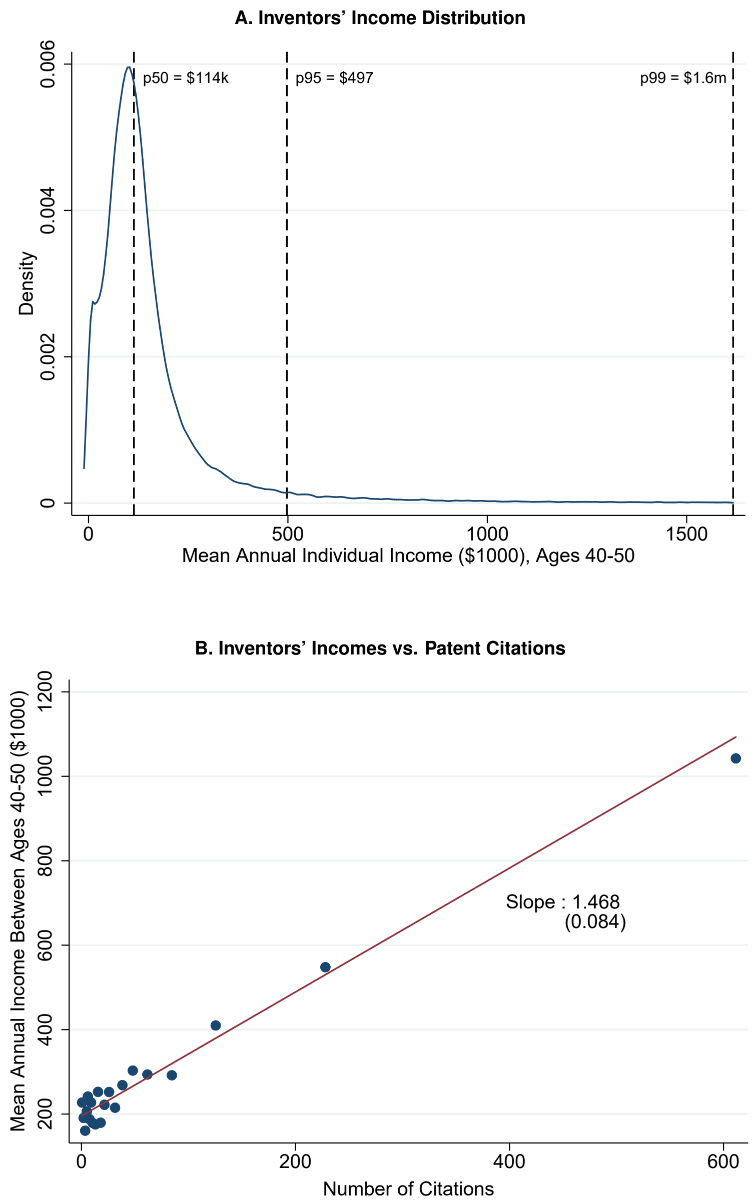
Notes: These figures present kernel densities of 3rd grade math test scores for children in the 1979-1985 birth cohorts who attended New York City public schools. Test scores, which are based on standardized tests administered at the district level, are normalized to have mean zero and standard deviation one by year and grade. In Panel A, we divide children into two groups based on whether their parents' incomes fall below the 80th percentile of the income distribution of parents' income in the New York City schools sample. Panel B compares boys and girls.

ONLINE APPENDIX FIGURE IV: Female Inventor Share and Gender Stereotype Adherence



Notes: This figure plots the share of inventors who are female vs. Pope and Syndor's (2010) gender stereotype adherence index by state. Female inventor shares are taken directly from Figure Xa; see notes to that figure for details. The stereotype adherence index is computed as $S = (N_{m,\text{math\&science}}/N_{f,\text{math\&science}} + N_{f,\text{reading}}/N_{m,\text{reading}})/2$, where $N_{g,s}$ denotes the number of students of gender $g \in \{m, f\}$ who score among the top 5% of students in their state in subject $s \in \{\text{math \& science, reading}\}$ in 8th grade. The index S measures the degree to which students adhere to the typical gender stereotype that boys do better at math/science and girls do better in reading; higher values represent greater adherence to this stereotype. The solid best-fit line is estimated using an unweighted OLS regression (slope = -4.49, standard error = 1.42).

ONLINE APPENDIX FIGURE V: Income of Inventors



Notes: Panel A plots a kernel density of the distribution of inventors' income, measured as mean annual income over ages 40-50 in 2012 dollars. Income is measured at the individual level and includes both labor and capital income. For scaling purposes, the top and bottom percentiles of the distribution are omitted. The dashed lines mark the median, 95th percentile, and 99th percentile of the distribution. In both panels, the sample consists of all individuals in our full inventors sample born between the ages of 1959-1962, for whom we see income at all ages between 40 and 50. Panel B presents a binned scatter plot of average annual income between ages 40 and 50 vs. the total number of citations an inventor obtains. For this panel, we further limit the sample to the 13,875 individuals who applied for a patent in 1996 to maximize the time horizon over which we can measure future citations. This plot is constructed by dividing citations into 21 bins and plotting mean income vs. mean citations within each bin. The first 19 bins include inventors in the first 19 ventiles (5% bins) of the citations distribution, while the last two bins plot the same relation for the 95th to 98th percentiles and the 99th percentile of the citation distribution. The best fit line and slope shown on the figure are estimated using an OLS regression on the 21 points, weighted by the number of inventors in each bin. The standard error of the slope estimate is reported in parentheses.