Access to Opportunity in the Sciences: Evidence from the Nobel Laureates*

Paul Novosad,[1†] Sam Asher,[2] Catriona Farquharson,[3] Eni Iljazi[4]

[1]Dartmouth College

[2]Imperial College London

[3]Princeton University

[4]UPenn Wharton

October 2024

Abstract

Unequal opportunity in the sciences reduces scientific contributions from the most talented individuals and limits the rate of human progress. We study unequal opportunity by collecting data on the childhood SES of Nobel laureates in the sciences. The average laureate grew up in an 87–90th percentile household. Access to opportunity doubled from 1901–2023, but remains highly unequal. Barriers are higher for women, but lower for Americans. Access to opportunity across countries is much less equal, and has barely improved at all. Cities with more intergenerational mobility produce more laureates from non-elite families, and more laureates overall.

Keywords: science of science, intergenerational mobility, access to opportunity

*This project was funded by Fast Grants for Innovations in STEM Talent Identification and Development. This paper benefited from excellent research assistance from Benjamin Levesque and Charlie Rudge. We are grateful for very helpful discussions with Kipp Bradford, James Feyrer, Douglas Staiger, and Heidi Williams.

"I am, somehow, less interested in the weight and convolutions of Einstein's brain than in the near certainty that people of equal talent have lived and died in cotton fields and sweatshops."

— Stephen Jay Gould

1 Introduction

Access to opportunity is unequally distributed: high ability people who grow up in difficult circumstances may not get the chance to develop their abilities and achieve their potential (Clark and Bono, 2016; Chetty et al., 2014; Zimmerman, 2019; Jia and Li, 2021; Bell et al., 2019). It is particularly bad for society when high ability people face barriers to success in the sciences, because scientific discoveries bring substantial benefits to all of humanity (Romer, 1990; Aghion and Howitt, 1992; Galor and Tsiddon, 1997; Griffith et al., 2004; Akcigit et al., 2017). If these barriers are large, then the pace of scientific progress could be accelerated substantially by improving access to opportunity.

Studying the childhood circumstances of scientists can shed light on the extent to which external barriers prevent people with high latent potential from success in the sciences. If talent is equally distributed, but scientists are disproportionately drawn from a subset of the population (say, defined by income, gender, or nationality), the implication is that a large number of high-ability people are not being discovered, and their talents are not maximally benefiting humankind.

In this paper, we propose a quantitative measure of unequal access to opportunity in the sciences, based on the distribution of childhood socioeconomic circumstances of top scientists. We focus on Nobel laureates — scientists who reached the peak of their respective disciplines and in virtually all cases produced discoveries of substantial and indisputable benefit to humankind. To measure childhood circumstances, we collected biographical data on the occupation of each laureate's father during the laureate's childhood. Occupation is a strong predictor of father income and education, and thus provides a measure of the laureate's socioeconomic circumstances during childhood. Occupational measures like this are widely used in historical studies on upward mobility when long panel data on individuals are unavailable (Duncan, 1966; Björklund and Jäntti, 1997; Long and Ferrie, 2013; Olivetti and Paserman, 2015; Song et al., 2020).

Specifically, based on their father's occupation, we assigned each laureate a 0–100 socioeconomic rank score using U.S. Census data, creating a parsimonious measure of childhood socioeconomic circumstances.[1] This rank score provides a natural benchmark for studying access to opportunity. In a world where access to opportunity is equal and outlier talent is randomly assigned, the average Nobel laureate would have a socioeconomic rank of 50. By contrast, if every laureate had a father who was a scientist, this would correspond to an income rank of 98 and an education rank of 99.9 — the ranks occupied by scientists in the historical income and education distributions — and would imply that the vast majority of people with latent scientific talent are not being discovered and nurtured to achieve their full scientific potential. The average rank score of laureates thus provides an easily interpretable measure of access to scientific opportunity that can be tracked over time.

This empirical setup motivates two questions. First, what is the average childhood socioeconomic rank of Nobel laureates, i.e. how equal is access to opportunity in the sciences? Second, how has this measure of access to opportunity changed over time — have human societies and institutions improved their ability to harness our collective talent over the 125 years since the first Nobel Prize was awarded?

We present six results. First, unsurprisingly, we show that Nobel laureates are drawn from elite families. The average laureate has a father at the 87th income percentile and the 90th education percentile. About 50–60% of laureates are from top 5% families.[2] Second, the distribution range from which laureates are drawn has approximately doubled since 1900, representing a substantial increase in access to opportunity at the very top of the sciences. The average winner was from a 92nd income percentile family in 1900, and is from an 85th income percentile family today — the mean education rank fell from 95 to 88 over the same period. There are no significant differences across scientific fields.

Third, we show that female laureates come from more elite families (91st income percentile, compared with 87th for men), suggesting that some of the barriers to advancement in the sciences

[1] We created separate ranks based on the father's income and the father's education (both predicted from occupation), and we analyze each separately. Predicting status from mother's occupation is less effective for reasons discussed below.

[2] The most common father occupations of laureates, in order, are business owners, physicians, professors, and engineers.

faced by women have been mitigated by family resources. Fourth, U.S.-born laureates come from lower SES ranks, suggesting higher access to scientific opportunity in the United States than in the rest of the world.

Fifth, restricting our sample to U.S.-born laureates, we show that laureate production by commuting zone (and also by county) is highly correlated with modern measures of intergenerational mobility. Using data from the Opportunity Atlas, we show that CZ upward mobility (p_{25} from Chetty et al. (2018)) is positively correlated with the production of Nobel laureates, as well as with the production of Nobel laureates from outside the top 90%. However, laureate production is *inversely* correlated with p_{75}, the expected income rank of children born at the 75th percentile. In other words, places where high-income children are likely to have high incomes (i.e. high p_{75}) produce *fewer* top scientists. This is somewhat surprising, because laureates tend to come from high income families. While we have not proved causality, the results are consistent with a story where more economic churn produces more superstar outcomes; better allocative efficiency or higher effort incentives could both explain this result.[3]

Finally, we extend our primary measure to include country of birth in addition to occupation. The rank results above are based strictly on father occupations; they ignore the fact that a child at the 80th percentile in the U.S. likely has better access to a scientific career than an 80th percentile child in India. We calculate global childhood SES ranks by ranking individuals in a synthetic global income distribution, which takes into account parent occupational earnings and national incomes.[4] The results highlight the substantial global inequality in the production of top scientists. The average laureate is at the 95th global income percentile. This rank has fallen by only a statistically insignificant 2 percentage points over the last 125 years — indicating little improvement in the world's ability to harness its top talent for science.

In Section 4, we discuss several alternate scenarios which would weaken the social cost implica-

[3]We also note that the Opportunity Atlas measures upward mobility for the 1980s birth cohorts, while the median laureate in our sample is born between in 1920 (IQR 1900–1940). There is limited evidence on the persistence of place effects on upward mobility.

[4]Specifically, we scale parent occupational earnings (which are based on U.S. data) by a country's GDP ratio with respect to the United States. We then re-rank all individuals globally, taking into account estimated national differences in occupation distributions.

tions of our findings, including: (i) people with outlier latent scientific talent could be producing in other socially valuable domains; (ii) talented people could be producing valuable science, but are going unrecognized; and (iii) high SES ranks could reflect genetic factors. While each of these mechanisms is plausible, we argue that they are unlikely to explain away the implied misallocation described here.

Unequal access to opportunity in the sciences is harmful not only because it is unfair, but because it can lead to a reduction in scientific discovery, indirectly harming all of humankind. The returns to research and discovery are large, as are the social returns to improving access to opportunity and the allocation of talent (Murphy et al., 1991; Jones and Williams, 2000; Bloom et al., 2019; Hsieh et al., 2019). Expanding the pool of talented people in the sciences also has positive externalities on the production of scientific research (Hoogendoorn et al., 2013; Freeman and Huang, 2015; Hunt and Gauthier-Loiselle, 2010; Gaulé and Piacentini, 2013).[5]

There has been significant prior research on the backgrounds and career paths of top scientists and inventors, though most of the work on scientists has focused on their career paths after high school rather than their childhoods (Akcigit et al., 2017; Holland, 1957; Zuckerman, 1967; Berry, 1981; Hargittai, 2002; Morgan et al., 2022). The study of parental occupations of scientists goes back as far as Galton (1874); our study is novel both in modernizing the data and in developing a parsimonious quantitative framework for examining changes in access to opportunity over time. Stansbury and Rodriguez (2024) notably study current relationships between career success in academia (and specifically economics in Stansbury and Schultz (2023)) and parental education, finding a substantial class gap, but does not examine changes over time. Several other papers study racial and gender gaps in academic advancement and recognition (Ginther et al., 2011, 2018; Sarsons et al., 2021; Hengel, 2022; Ross et al., 2022; Koffi et al., 2024).

Our paper is related to concurrent work by Airoldi and Moser (2024), who study the parental backgrounds of top U.S. scientists in the 1920s, finding that high SES predicts entry into the sciences and continues to drive professional success throughout the career cycle. Our work is also related to

[5]Increasing representation among scientists is also independently important; for example, female inventors are more likely to produce patents related to women's health (Koning et al., 2021).

Bell et al. (2019), who examine the childhood socioeconomic ranks of U.S. inventors in a similar quantitative framework, and to work examining careers of women in science (Kim and Moser, 2021). Research on immigration and science suggests significant returns to increasing access to scientific opportunity for talented foreigners (Bloom et al., 2019; Hunt and Gauthier-Loiselle, 2010; Gaulé and Piacentini, 2013; Agarwal et al., 2023). A related literature studies the networked relationships between scientists who receive major prizes (Ma and Uzzi, 2018; Jin et al., 2021).

Finally, our paper relates to a wide body of work studying intergenerational mobility and access to opportunity (Chetty et al., 2014; Solon, 1992). Due to limitations in the historical data, studies of changes in intergenerational mobility over time are comparatively uncommon (Olivetti and Paserman, 2015; Song et al., 2020; Chetty et al., 2017). Our data on Nobel laureates provides a rare 125-year time series, allowing us to measure how access to opportunity (albeit in only one important domain) has changed over the long run. Our work on counties and commuting zones also relates to the wide literature showing that neighborhood characteristics matter for individual outcomes (Chetty et al., 2016; Chetty and Hendren, 2018a).

2 Biographical Data and Measurement of Socioeconomic Rank

We collected biographical information on prize winners from 1901–2023 from a range of historical sources, covering 739 winners in Chemistry, Physics, Medicine, and Economics.[6] We exclude the prizes for Peace and for Literature, as accomplishments in these domains are more subjective and are sometimes explicitly awarded to individuals who faced challenging childhood circumstances. For each winner, we collected demographic information and parents' occupations.[7] We supplemented the data by contacting living winners directly to request information about their parents. In total, we were able to identify fathers' occupations for 715 out of the 739 laureates, but mothers for only 181; our analysis therefore focuses on father occupation as a measure of childhood socioeconomic

[6]The Economics prize was established only in 1968; results are very similar if the field is excluded.

[7]Biographical information on winners was collected from biographies and interviews on the official web site of the Nobel Prize, other interviews with Nobel laureates, entries from *Encyclopedia Britannica*, profiles published by winners' academic institutions, and articles describing early life found in various other biographical sources.

status.[8]

Using father occupation as a measure of childhood socioeconomic status is a common approach in the historical intergenerational mobility literature (Long and Ferrie, 2013; Song et al., 2020; Long and Ferrie, 2007). Father occupation is a good predictor of other measures of socioeconomic status, and is often the only such variable available in historical sources. We classified the relative status of occupations using census data on average income and education levels of occupation holders. We use the IPUMS data from the 1850–1990 U.S. population censuses to calculate occupational earnings and education levels covering the birth years of the majority of winners. For the 12 winners born before 1850, we used occupational data from 1850. Years of education and income are recorded in the population census beginning in 1940 and 1950 respectively; we use 1940 measures for all earlier periods.[9] We mapped fathers' occupations to the 269 unique occupational categories classified in the 1950 census; where there was ambiguity over occupation classification, we tested robustness to alternate coding approaches.[10]

We calculated income and education ranks for each occupation category in each decade, and matched them to prize winners' birth decades and fathers' occupations.[11] These ranks reflect the average income or education rank that would have been held by an adult male in the given occupation in the birth year of the prize winner. Income and educational ranks are highly but imperfectly correlated; some example outliers are members of the clergy and (to a lesser extent) teachers (high education but low income ranks), and tradesmen (high income but low education ranks).

Some occupations have very wide income distributions in the data; for instance, in 1850, farmers represented 65% of the American workforce, some of whom were very rich and others very poor. Entrepreneurs are also difficult to classify; the biographical data tells us a parent is a business

[8]Mother's occupation is also less predictive of socioeconomic status than fathers, as many educated women were out of the labor force, particularly in the earliest years of our sample.

[9]The census recorded information on literacy from 1850–1930, but we elected not to use these data, given their substantial inaccuracies (see, for example, Collins and Margo (2006)).

[10]Appendix B.2 describes additional details on the calculation of occupational earnings and education levels.

[11]While the earnings and education level associated with each occupation are held constant for all years before 1940, the ranks change over time because of changes in the distribution of occupations in the country.

owner, but is often vague about the scale of the business. To the extent that prize winners are disproportionately born to the highest SES individuals within an occupation, our measure would bias our rank estimates downward, raising the observed equality of opportunity. We therefore show that our results are robust to excluding or alternately ranking these difficult-to-rank occupations.

Occupational rank data are not available for all countries in which prize winners were born, so our estimates are based on the U.S. occupational rank distribution. However, we adjusted for the different occupational distributions in less-developed countries by matching country-decade pairs to the U.S. decade with the most similar level of development. For example, in 1960, a professor in India would have a higher relative socioeconomic rank in India than a professor would have in the U.S.. We would therefore classify the Indian professor using the U.S. rank distribution from 1850, the year in the U.S. data with the most similar per capita GDP to India in 1960 — thus assigning them a higher socioeconomic rank within their country.

We describe these as *national* or *domestic* rank measures, as they describe the relative status of winners' parents within their own countries. These measures do not reflect differences in opportunity across countries, which may be significant.

To address income differences across countries, we additionally construct a synthetic global income distribution and rank parents within it. For each country-decade pair, we create a synthetic distribution of individuals, reflecting the occupation distribution in the U.S. decade with the nearest per capita GDP, as above. Next, we assign each person an occupational income which is equal to the 1950 U.S. census income for that occupation, multiplied by the proportional GDP difference between that country and the United States, in the given decade. We then rerank all individuals in this synthetic global income distribution, creating a new global income rank for each country-decade-occupation group. For example, a tailor in 1950 has a U.S. occupational income rank of 55 — their position in the U.S. national income distribution. According to our global measure, a U.S. tailor in 1950 would be at the 97th percentile globally, reflecting high U.S. incomes compared with the rest of the world. Conversely, we would classify an Indian tailor in 1950 at the 46th global percentile, reflecting India's relative poverty in 1950. See Appendix B.3 for additional details and examples.

3 Results

3.1 Results: SES of Nobel Laureates from 1901–2023

The analysis dataset describes 715 winners from 44 countries over a period of 123 years. About 4% of winners are women, 35% are U.S.-born, and an additional 12% are from the United Kingdom.[12]

Figure 1 plots the distribution of national father income percentiles. 67% of winners have fathers from above the 90^{th} income percentile in their birth country; 75% have fathers above the 90^{th} education percentile. In both cases, most winners are from the top 5%. The mean laureate has a father at the 87th income percentile, and the 90th education percentile. If talent and opportunity were equally distributed, the average winner's parent would be at the 50th percentile. Under these assumptions, our estimates would suggest that over three quarters of children born with top latent scientific talent did not receive the complementary parental and social inputs necessary to bring their talent to maximal fruition.

Figure 2 shows the ten most frequent occupations held by the fathers of laureates, along with the occupation's population share. Business owners, physicians, professors, and scientists feature most prominently. Relative to their occupational shares in the census data, professors, physicians and scientists are over-represented by a factor of 100 or more, while business owners are over-represented by a factor of 2.5. The frequency distribution of parental occupations is largely similar across the different prize categories (Appendix Table A1).

We next examine how the relative childhood status of prize winners has changed over time. Figure 3 shows the average ranks of winners in 5-year bins since the first Nobel Prize, along with a linear estimation of the time trend. An equivalent regression estimation of the time trend is displayed in Appendix Table A2. Every 100 years, on average, there is a 6 percentage point ($p < 0.01$) decline in the average national income and education rank of laureates' fathers. In other words, recent prize

[12]There are six laureates whose fathers were also Nobel laureates — one of whom (William Lawrence Bragg) shared the prize with his father (William Henry Bragg) — and one winner (Irène Joliot-Curie), whose parents were both laureates.

winners are being drawn from a broader set of socioeconomic backgrounds than earlier winners. Based on the mean, the pool of winners is twice as socioeconomically broad as it was in 1901. The pace of change is perhaps slow: at the current rate, it would be another 600 years before Nobel Prize winners had similar family backgrounds to the population average.[13]

3.2 Results: Heterogeneity by Gender and Region

We next examine how the distribution of parent ranks depends on various geographic and demographic factors. If potential scientists from a given subgroup (e.g. women) face greater social barriers to scientific success, we expect that winners in that subgroup will come from higher status parents. This reflects the fact that they require greater parental inputs to make up for the extra barriers to progress that they face.

3.2.1 Results: Inequality by Gender

Female winners come from more elite families than male winners; the average female laureate has a parent at the 91st percentile, compared to the 87th percentile for the average male winner (Table 1, Panel A). Figure 4 compares the full distribution of parent education and income ranks for both genders; the excess mass at the highest ranks for women is evident in both density functions.

By every measure (mean parent rank, share of winners in the top 10% of SES, and share of winners in the top 5% SES), female laureates come from more elite families than male laureates (Table 1, Panel A). The differences are not all statistically significant, owing to the small sample of women (only 28 laureates are female).[14] Appendix Figure A2 shows the most common father occupations among the set of female laureates. The main difference in occupational category is that female laureates are less likely than males to have fathers who were business owners or managers, but note the small sample qualifications above.

[13] Appendix Table A3 shows that these estimates are robust to excluding occupational categories with more subjective rank classification, like aristocrats, farmers, business owners, and others.

[14] The distribution of women's SES ranks is more elite in education than it is in income; many female winners are in income percentiles 90–95 and education percentile 95–100. But the small sample of women suggests caution in interpreting this finding. Appendix Table A5 shows that these estimates are highly robust to excluding occupational categories with more subjective rank classifications.

Appendix Figure A1 plots the proportion of winners below the 90th and 95th income and education percentiles, by decade. While the graphs are consistent with the distribution of male and female ranks becoming more similar, the sample of women is too small to make a strong statistical claim about changes over time.[15]

The raw share of female winners (4%) already makes it evident that women have faced greater barriers to success in the sciences. Our analysis shows that high socioeconomic status can counteract some of these barriers, but it means that elite women in the sciences are drawn from an even smaller socioeconomic pool than men, implying substantial misallocation.

3.2.2 Results: Inequality by Field

Table 1B shows differences in the childhood socioeconomic status of Nobel laureates by prize category. The average ranks of winners are notably similar across fields, with all fields within two rank points of another on both mean measures, and no statistically significant differences.[16]

3.2.3 Results: Inequality by World Region

Table 1C shows the regional pattern of winners' ranks. If some regions are better at nurturing scientific talent and creating opportunities for lower-income individuals, prize winners from these regions could be expected to come from relatively more humble backgrounds. The sample of countries outside of the U.S. and Europe is limited, but we nevertheless find some notable differences. U.S.-born winners tend to come from less elite backgrounds than winners born in Europe and the rest of the world, suggesting more equal access to opportunity in the sciences in the United States, at least in the century before 1960 when most of our winners grew up. The results are consistent whether we look at family education or income ranks, and the regional differences are highly statistically significant ($p < 0.01$). Appendix Figure A3 shows the average parental rank of U.S.-born vs. non-U.S.-born winners over time, showing the consistent pattern of U.S. winners coming

[15] In Appendix Table A7, we regress a rank measure on a female indicator, and an interaction between the female indicator and a post-2000 indicator. The interaction is consistently negative, but only statistically significant in one out of six specifications, and is sensitive to alternate choices about coding specific laureates.

[16] Appendix Table A8 shows linear trends estimated separately for each field.

from less elite backgrounds.[17]

3.3 Results: Local Access to Opportunity and Success in the Sciences

In this section, we focus on U.S.-born winners, and ask whether U.S. cities characterized by relatively equal access to economic opportunity have a higher propensity to produce top scientists, and particularly whether they tend to produce more top scientists from non-elite backgrounds. We use two commuting-zone-level measures of access to opportunity from the Opportunity Atlas (Chetty et al., 2018) to carry out these analyses. *Upward mobility* (p_{25}) is defined as the average income percentile rank (in the national income distribution) of a child born in a given commuting zone to a family at the 25th national income percentile (Chetty et al., 2014). Commuting zones (CZs) with low p_{25} are places where children born into poverty are likely to stay in poverty. We also examine p_{75}, which is the average income rank of a child born to a 75th percentile family. This is a measure of the persistence of high economic status—an inverse measure of downward mobility. CZs with high values of p_{75} are places where high income is persistent and the rich are unlikely to lose relative status across generations. We use the descriptive mobility estimates from Chetty and Hendren (2018a); see Appendix B.4 for additional details.

Our outcome variable is a measure of the extent to which a CZ produces prize winners. Our preferred measures are binary indicators of whether a CZ produced any Nobel laureate at all, or whether it produced any laureate from the bottom X% of the SES distribution. We prefer these to continuous measures of the counts of winners, as the latter are heavily skewed by a handful of large cities with a huge share of winners (*e.g.* New York (65), Boston (21), and Chicago (15)).

Table 2 shows estimates from regressions of the prize measures on the mobility measures p_{25} and p_{75}.[18] We include a polynomial population control because large cities are more likely to produce winners just by virtue of having more people.

[17] Appendix Table A9 shows regression estimates from just the United States against the rest of the world. Appendix Table A10 shows the list of countries by classification region.

[18] Appendix Table A11 shows analogous estimates using the causal measures of p_{25} and p_{75}, which are highly similar in significance and magnitude. Appendix Tables A12 and A13 repeat this exercise at the county level. The county-level effects are broadly similar to the CZ results, but slightly smaller and less precise since only 4% of counties have produced any Nobel laureates, compared to 13% of CZs. See Appendix B.4 for additional details.

The results show that places with high equality of opportunity are substantially more likely to produce Nobel laureates. Moving from the lowest opportunity to the highest opportunity commuting zone (a 43 rank point change in p_{25}) is associated with a 46 percentage point increase in the probability that a place produces at least one Nobel laureate (Column 2), a 17–33 percentage point increase in the probability of producing a laureate from a family outside of the top 10% (Columns 3–4), and a 25–34 percentage point increase from a family outside of the top 5% (Columns 5–6). Places where poor people have opportunities to move up the income ladder are better at nurturing scientific talent in *all* social classes.

Notably, we find statistically and economically significant effects for p_{75} with the opposite sign. This may at first seem surprising, since prize winners overwhelmingly come from top 25% families. In fact, high status families tend to produce exceptional scientists in exactly the places where high incomes are *not* guaranteed for the children of the rich.

Places with high persistence of high income (i.e., high p_{75}) may be places where the rich are relatively entrenched in their social positions and institutions are less meritocratic; there is thus less ability-based sorting into upper-middle class occupations and worse matching of talent. For instance, in a context where only the children of scientists can become scientists (and similarly for other occupations), then children of scientists could be expected to have relatively high incomes (i.e., high p_{75}). But children with top scientific potential from non-scientist families might not receive the inputs and access necessary to succeed in the sciences. Alternately, talented individuals might work harder in a social environment where their economic success is not guaranteed. Further exploration of the mechanisms that link local upward and downward mobility to local scientific achievement could be a worthwhile direction for future work.

Needless to say, these results are only suggestive. We have not demonstrated a causal mechanism; it is possible that families likely to produce top scientists choose to move to places with high $p25$ and low p_{75}. We also note that the Opportunity Atlas measures are based on birth cohorts from the early 1980s, while Nobel laureates are from much older cohorts. CZ and county characteristics may certainly have changed between the two samples; but classical measurement error would bias these estimates toward zero. Note finally that these estimates are informative in a predictive sense; places with high intergenerational mobility do produce more scientists, even if we cannot prove that upward mobility is a mechanism for this outcome.

3.4 Results: Accounting for Income Differences Across Countries

Our measures thus far have captured the dimensions of inequality that are correlated with parent occupation, but have ignored the substantial differences in opportunity across countries. As described above, we reproduce our primary estimates of the average socioeconomic rank of Nobel laureates, but this time using global income rank estimates which take into account both the father occupation and the GDP per capita of each winner's childhood country.[19] Figure 5 shows the average global income rank of Nobel laureates from 1901–2023, with the national income rank estimates from above as a benchmark. The results are stark. The average global rank of a prize winner is 95; this has improved only by a statistically insignificant two rank points over the 122 years since 1901. The estimates imply that the vast majority of the world's best scientific talents are growing up in contexts that prevent them from achieving their potential and sharing it with humanity, and that little progress has been made to close the cross-country component of this gap.

4 Discussion

Our evidence suggests that there is a large number of "missing scientists" — individuals who could have produced important scientific discoveries, but did not receive the complementary inputs required over their lives to do so. Precisely determining the social cost of missing these scientists would require us to understand what these most talented people do, if not achieve maximal success in the sciences. In this section, we examine some factors that could influence our thinking about the social cost of having top scientists drawn predominantly from elite families.

Are missing scientists producing equally valuable outcomes in non-scientific domains, such as

[19]We define the winner's country as the country where they permanently resided at age 6. See Appendix B.1 for additional details.

business or medical practice? There are three reasons to think that their social contributions in other fields are likely to be smaller than they would be in the sciences. First, the social benefits of research in the basic sciences are plausibly among the highest out of all activities, because the externalities of research are so large (Jones and Williams, 1998; Bloom et al., 2013; Gross and Sampat, 2023). Second, talent is multi-dimensional; the individuals who would be most successful in the sciences would be unlikely to find equal levels of success in other domains (Murphy et al., 1991). Third, the unequal barriers that prevent advancement in the sciences also exist in other domains. For example, authors of highly-cited patents are ten times more likely to come from the top 5% of the parent income distribution than from the bottom 50% (Bell et al., 2019); nationality and gender play a role in this inequality as well. Statistics are similar for doctors, CEOs of large companies, and politicians (Dal Bó et al., 2017; Thompson et al., 2019). The evidence suggests that talented individuals who face social barriers to success in the sciences are not likely to find easy paths in other domains either.

An alternative hypothesis is that missing scientists are in fact generating important scientific discoveries, and are simply not being recognized for their work. For example, the physicist Lise Meitner was nominated for the Nobel Prize more than 40 times for her work on nuclear fission but was overlooked when the prize was given to her colleague Otto Hahn. If all of our "missing scientists" had stories like these, then the inequality of opportunity identified in this paper would have less consequence for scientific discovery. But it is unlikely that most of our missing scientists are in this category. To reach the top of the sciences, an individual must pass through many filters, each of which may disadvantage people from less affluent backgrounds. Entry into elite PhD programs and postdocs, grants, and early career awards are all key ingredients of success, and all of these are based on evaluation of one's work by peer scientists, like the Nobel Prize itself. It is therefore unlikely that the unequal outcomes we describe here are driven entirely by unequal selection at the final stage of one's scientific career; it is more likely the culmination of multiple inequalities that have accumulated over the full life cycle. But in this case, we would not expect a mass of star scientists having achieved everything *except* the Nobel Prize; instead, they are being lost at every stage of the pipeline, and thus unlikely to be producing discoveries at their full potential.

Finally, we note the omission of genetic factors in our analysis. If talent is genetic, and high-talent people are sorted into occupations with high income and education, then this can explain some of the correlation between parent occupation and scientific success. While the unobserved genetic factor undoubtedly puts a qualification on the idea that prize winners would come from 50th percentile households in an equal opportunity world, we think this is a limited concern for three reasons.

First, the genetics of extreme intelligence are unknown, and modern science in particular requires lab and personnel management and many other skills besides raw intelligence, which could limit the role of genetic factors. Second, the exclusionary mechanisms that have prevented access to scientific opportunity for women and minorities are widely known and were likely even larger in the era covered by our analysis. Bell et al. (2019) find that primary school test scores (a partial proxy for genetic talent) can explain less than a third of the SES inequality in patent authorship. Further, the cross-country differences in access to scientific opportunity — a huge share of the total inequality — cannot plausibly be explained by genetic effects. Third, our time series results demonstrate substantial improvement in equitable access to scientific opportunity. The justifications in the past of inequality based on arguments from genetics should make us cautious about accepting these arguments without clear evidence today. Determining the genetic basis of scientific talent is a challenging task that is well beyond the scope of our paper, but of would nevertheless be helpful in interpreting our results.

5 Conclusion

Talent is widely distributed; when opportunity is unequal, great talents are under-nourished and potential scientific discoveries are lost. Nobel laureates in the sciences overwhelmingly originate from the top tiers of national income distributions, and even moreso from global income distributions. There has been some progress toward broadening access to opportunity in the United States and in Europe, but comparatively little progress at bridging global income divides.

Our results suggest that there is a vast reservoir of untapped scientific talent in lower-income countries, who lack the means to cultivate their abilities, limiting human potential in the sciences. The small number of female winners and their elite backgrounds also highlight a substantial untapped potential. More inclusive access to opportunity in the sciences is valuable for its own sake, but could also unleash new avenues of innovation, growth, and advancement for humankind.

References

Agarwal, R., Ganguli, I., Gaulé, P., and Smith, G. (2023). Why u.s. immigration matters for the global advancement of science. Research Policy, 52(1):104659.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. Econometrica, 60(2):323–351.

Airoldi, A. and Moser, P. (2024). Inequality in science: Who becomes a star? Working Paper.

Akcigit, U., Grigsby, J., and Nicholas, T. (2017). The rise of american ingenuity: Innovation and inventors of the golden age. NBER Working Paper No. 29422.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., and Van Reenen, J. (2019). Who becomes an inventor in america? the importance of exposure to innovation. The Quarterly Journal of Economics, 134(2):647–713.

Berry, C. (1981). The nobel scientists and the origins of scientific achievement. The British Journal of Sociology, 32(3):381–391.

Björklund, A. and Jäntti, M. (1997). Intergenerational income mobility in sweden compared to the united states. The American Economic Review, 87(5):1009–1018.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. Econometrica, 81(4):1347–1393.

Bloom, N., Van Reenen, J., and Williams, H. (2019). A toolkit of policies to promote innovation. Journal of Economic Perspectives, 33(3):163–184.

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., and Porter, S. R. (2018). The opportunity atlas: Mapping the childhood roots of social mobility.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., and Narang, J. (2017). The fading american dream: Trends in absolute income mobility since 1940. Science, 356(6336):398–406.

Chetty, R. and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. The Quarterly Journal of Economics, 133(3):1107–1162.

Chetty, R. and Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. The Quarterly Journal of Economics, 133(3):1163–1228.

Chetty, R., Hendren, N., and Katz, L. F. (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., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. The Quarterly Journal of Economics, 129(4):1553–1623.

Clark, D. and Bono, E. D. (2016). The long-run effects of attending an elite school: Evidence from the united kingdom. American Economic Journal: Applied Economics, 8(1):150–176.

Collins, W. J. and Margo, R. A. (2006). Chapter 3 historical perspectives on racial differences in schooling in the united states. volume 1 of Handbook of the Economics of Education, pages 107–154. Elsevier.

Dal Bó, E., Finan, F., Folke, O., Persson, T., and Rickne, J. (2017). Who Becomes a Politician? Quarterly Journal of Economics, 132(4):1877–1914.

Duncan, O. D. (1966). Methodological issues in the analysis of social mobility. In Social Structure and Mobility in Economic Development. Aldine, Chicago.

Fariss, C. J., Anders, T., Markowitz, J. N., and Barnum, M. (2022). New estimates of over 500 years of historic gdp and population data. Journal of Conflict Resolution, 66(3):553–591.

Freeman, R. B. and Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the united states. Journal of Labor Economics, 33(S1):S289–S318.

Galor, O. and Tsiddon, D. (1997). Technological progress, mobility, and economic growth. The American Economic Review, 87(3):363–382.

Galton, F. (1874). English Men of Science: Their Nature and Nurture. Routledge.

Gaulé, P. and Piacentini, M. (2013). Chinese graduate students and u.s. scientific productivity. The Review of Economics and Statistics, 95(2):698–701.

Ginther, D. K., Basner, J., Jensen, U., Schnell, J., Kington, R., and Schaffer, W. T. (2018). Publications as predictors of racial and ethnic differences in nih research awards. PLOS ONE, 13(11).

Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., and Kington, R. (2011). Race, ethnicity, and nih research awards. Science, 333(6045):1015–1019.

Griffith, R., Redding, S., and Reenen, J. V. (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. The Review of Economics and Statistics, 86(4):883–895.

Gross, D. P. and Sampat, B. N. (2023). America, jump-started: World war ii r&d and the takeoff of the us innovation system. American Economic Review, 113(12):3323–3356.

Hargittai, I. (2002). The Road to Stockholm: Nobel Prizes, Science, and Scientists. Chemical Heritage Foundation.

Hengel, E. (2022). Publishing while female: are women held to higher standards? evidence from peer review. The Economic Journal, 132(648):2951–2991.

Holland, J. L. (1957). Undergraduate origins of american scientists. Science, 126(3271):433–437.

Hoogendoorn, S., Oosterbeek, H., and van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. Management Science, 59(7):1514–1528.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and u.s. economic growth. Econometrica, 87(5):1439–1474.

Hunt, J. and Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? American Economic Journal: Macroeconomics, 2(2):31–56.

Jia, R. and Li, H. (2021). Just Above the Exam Cutoff Score: Elite College Admission and Wages in China. Journal of Public Economics, 196(C).

Jin, C., Ma, Y., and Uzzi, B. (2021). Scientific prizes and the extraordinary growth of scientific topics. Nature Communications, 12(1):5619.

Jones, C. I. and Williams, J. C. (1998). Measuring the social return to r&d. The Quarterly Journal of Economics, 113(4):1119–1135.

Jones, C. I. and Williams, J. C. (2000). Too much of a good thing? the economics of investment in r&d. Journal of Economic Growth, 5(1):65–85.

Kim, S. D. and Moser, P. (2021). Women in science. lessons from the baby boom. Working Paper NBER Working Paper No. 29436, National Bureau of Economic Research.

Koffi, M., Pongou, R., and Wantchekon, L. (2024). Racial inequality and publication in economics. AEA Papers and Proceedings, 114:300–304.

Koning, R., Samila, S., and Ferguson, J.-P. (2021). Who do we invent for? patents by women focus more on womenâs health, but few women get to invent. Science, 372(6548):1345–1348.

Long, J. and Ferrie, J. (2007). The Path to Convergence: Intergenerational Occupational Mobility in Britain and the US in Three Eras. The Economic Journal, 117(519).

Long, J. and Ferrie, J. (2013). Intergenerational occupational mobility in great britain and the united states since 1850. American Economic Review, 103(4):1109–1137.

Ma, Y. and Uzzi, B. (2018). Scientific prize network predicts who pushes the boundaries of science. Proceedings of the National Academy of Sciences, 115(50):12608–12615.

Morgan, A. C., LaBerge, N., Larremore, D. B., Galesic, M., Brand, J. E., and Clauset, A. (2022). Socioeconomic roots of academic faculty. Nature Human Behaviour, 6(12):1625–1633.

Murphy, K. M., Shleifer, A., and Vishny, R. W. (1991). The allocation of talent: Implications for growth. The Quarterly Journal of Economics, 106(2):503–530.

Olivetti, C. and Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940. American Economic Review, 105(8):2695–2724.

Romer, P. M. (1990). Endogenous technological change. Journal of Political Economy, 98(5, Part 2):S71–S102.

Ross, M. B., Glennon, B. M., Murciano-Goroff, R., Berkes, E. G., Weinberg, B. A., and Lane, J. I. (2022). Women are credited less in science than men. Nature, 608(7921):135–145.

Sarsons, H., Gërxhani, K., Reuben, E., and Schram, A. (2021). Gender differences in recognition for group work. Journal of Political Economy, 129(1):101–147.

Solon, G. (1992). Intergenerational income mobility in the united states. The American Economic Review, 82(3):393–408.

Song, X., Massey, C. G., Rolf, K. A., Ferrie, J. P., Rothbaum, J. L., and Xie, Y. (2020). Long-term decline in intergenerational mobility in the united states since the 1850s. Proceedings of the National Academy of Sciences, 117(1):251–258.

Stansbury, A. and Rodriguez, K. (2024). The class gap in career progression: Evidence from us academia. Technical report. Working Paper.

Stansbury, A. and Schultz, R. (2023). The class gap in career progression: Evidence from us academia. Journal of Economic Perspectives, 37(4):207–230.

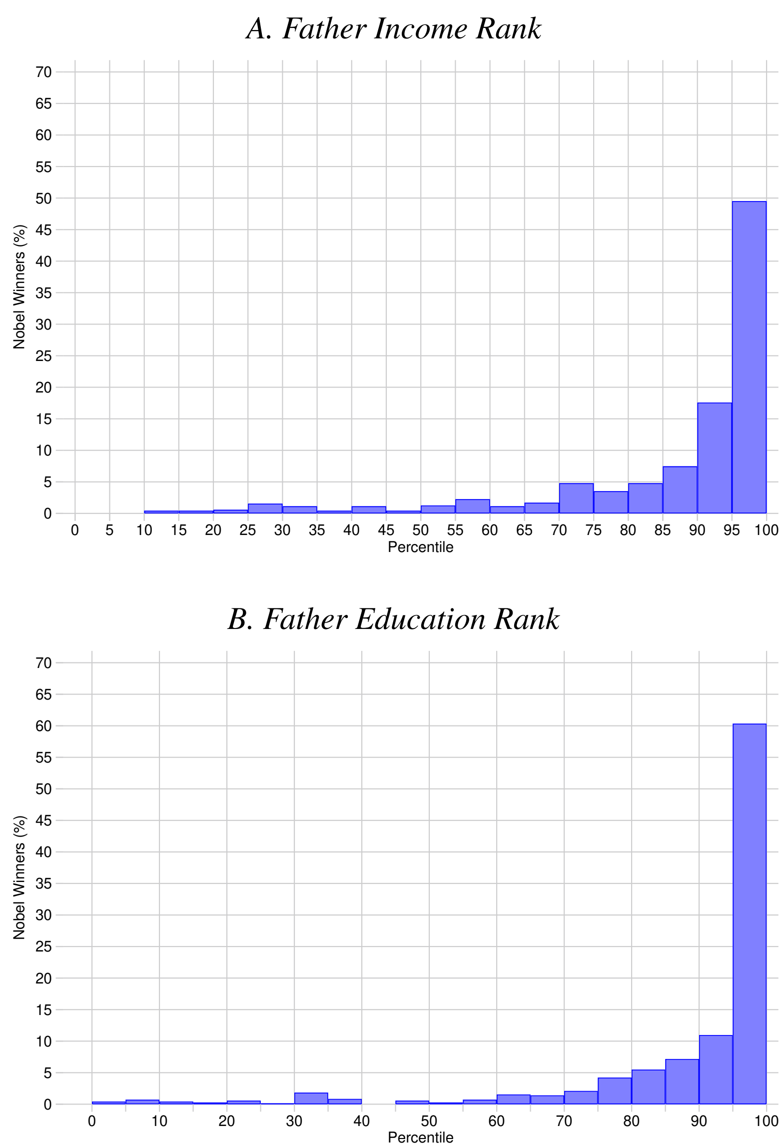
Thompson, D. M., Feigenbaum, J. J., Hall, A. B., and Yoder, J. (2019). Who becomes a member of congress? evidence from de-anonymized census data.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. American Economic Review, 109(1):1–47.

Zuckerman, H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. American Sociological Review, 32(3):391–403.

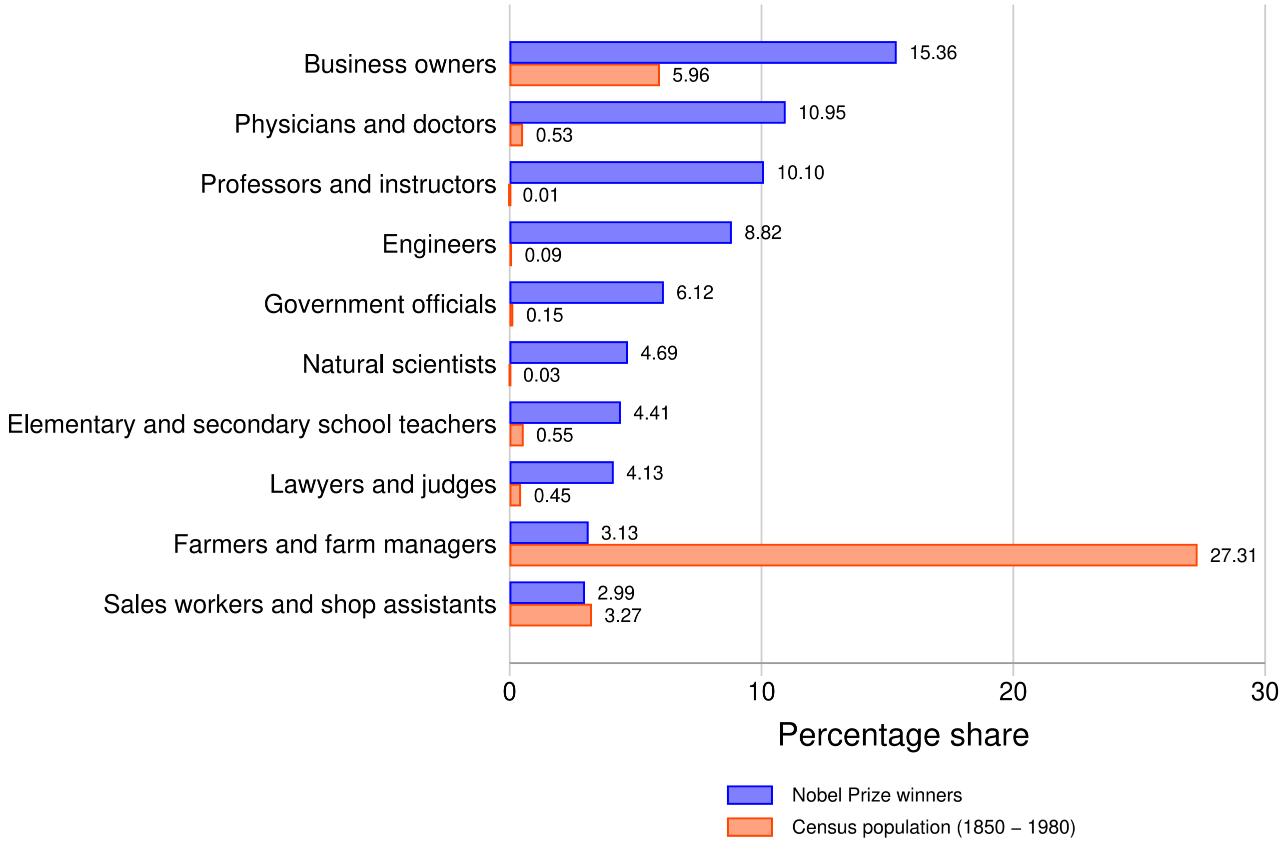
Figures and Tables

Figure 1: Distribution of Socioeconomic Ranks of Fathers of Nobel Laureates



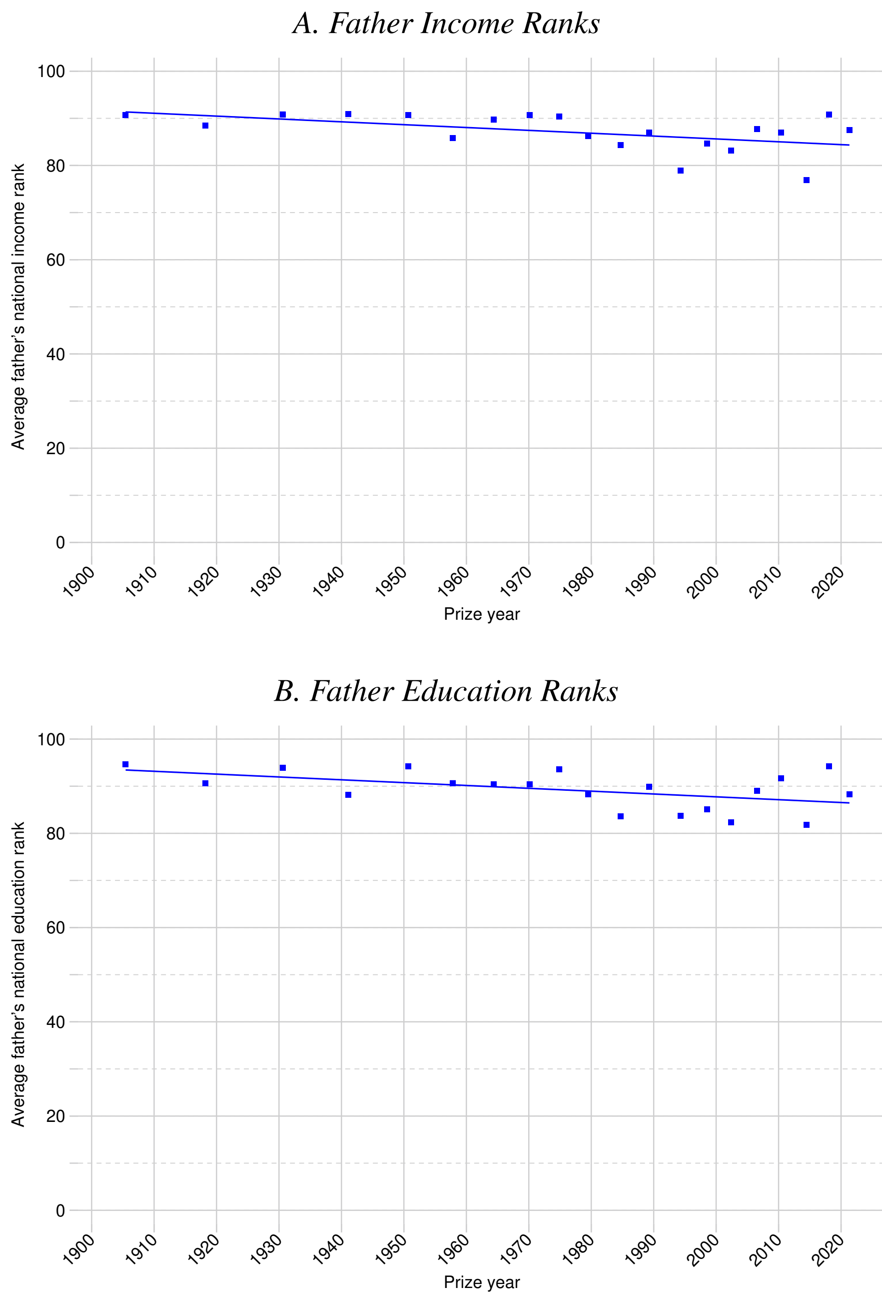
Notes: The graph shows the frequency distribution of national father income and education percentiles (bin size = 5) for all Nobel laureates in the sciences in our sample. Father income and education percentiles were predicted from father occupations, using data from the 1940 and 1950 U.S. Population Censuses.

Figure 2: Most Frequent Father Occupations among Nobel Laureates



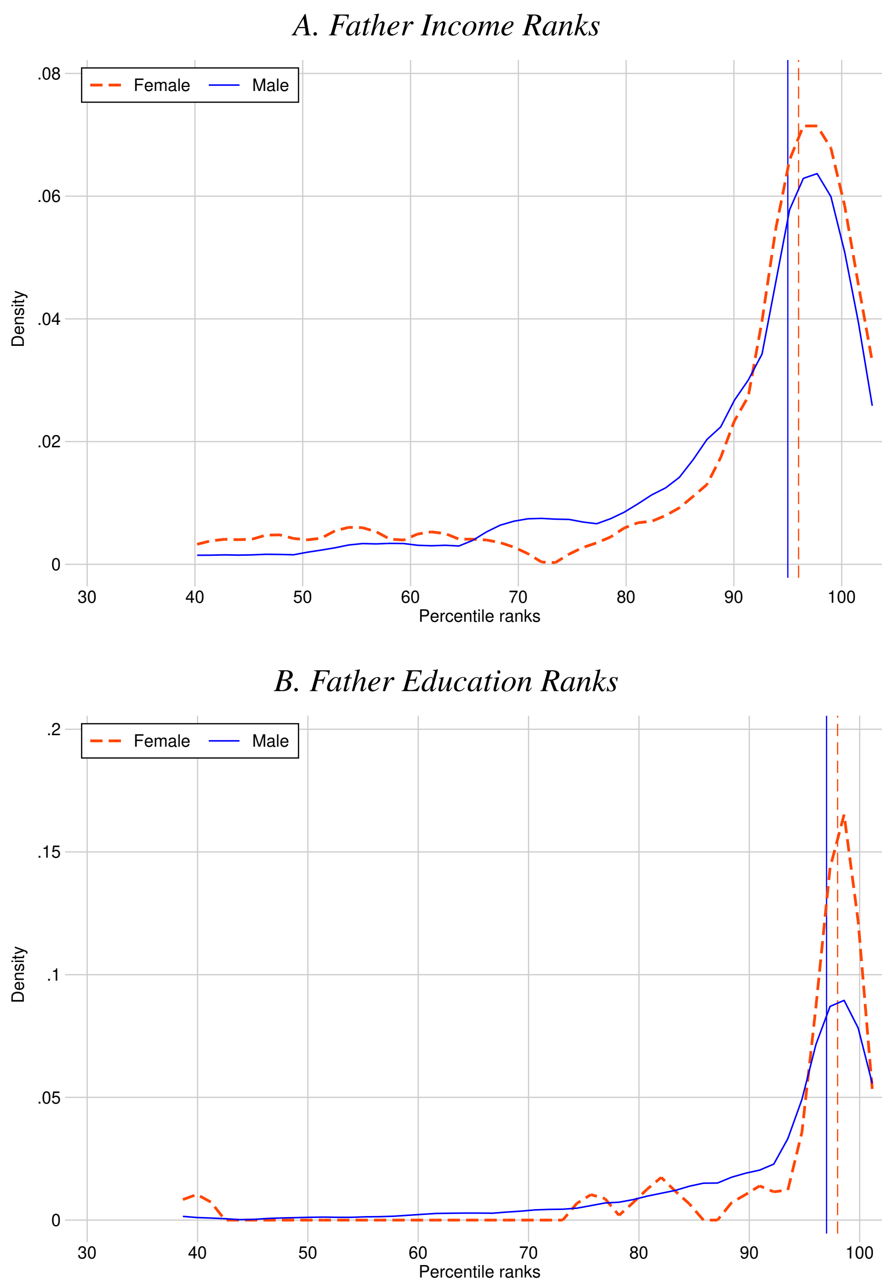
Notes: The graph shows the distribution of father occupations among Nobel laureates in the sciences (in blue) and the general population (in orange). Occupational shares are averages from the U.S. Population Census between 1850 and 1980. The laureate sample covers the full set of prizes from 1901–2023.

Figure 3: Average Socioeconomic Ranks of Fathers of Nobel Laureates



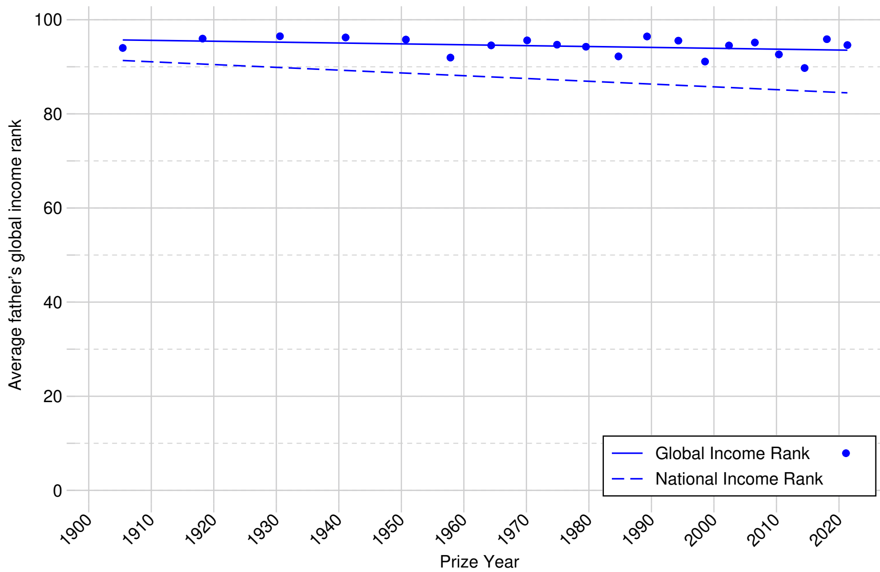
Notes: The graph shows binned scatterplots of national father's income and education rankings for Nobel laureates in the sciences from 1901–2023. Father income and education percentiles were predicted from father occupations, using data from the 1940 and 1950 U.S. Population Censuses. The fitted lines are estimated using ordinary least squares (OLS).

Figure 4: Average Socioeconomic Rank of Fathers of Nobel Laureates, by Laureate Gender



Notes: The graphs show the kernel density functions of national socioeconomic rankings of fathers of all Nobel laureates in the sciences, by laureate gender. The vertical lines represents the median percentile within each gender. Father income (Panel A) and education (Panel B) percentiles were predicted from father occupations, using data from the 1940 and 1950 U.S. Population Censuses.

Figure 5: Average Global Income Ranks of Fathers of Nobel Laureates



Notes: The graph shows a binned scatterplot of the average global income percentile of the fathers of Nobel laureates, over time. Incomes are constructed from father occupation data from the U.S. Census, and then adjusted for country occupation distributions and country GDP, as discussed in the data section. The national income ranks from Figure 3 are included (dashed line) as a reference. The fitted lines are estimated using ordinary least squares (OLS) regressions.

Table 1: Heterogeneity in Socioeconomic Rank of Fathers of Nobel Laureates

	Average Rank		**Top 10% Indicator**		**Top 5% Indicator**		
	National Income	National Education	National Income	National Education	National Income	National Education	*N*
Full Sample	87.081 (0.677)	89.187 (0.679)	0.671 (0.018)	0.713 (0.017)	0.495 (0.019)	0.603 (0.018)	711
Panel A: Gender							
Male	86.977 (0.695)	89.004 (0.701)	0.668 (0.018)	0.709 (0.017)	0.492 (0.019)	0.596 (0.019)	683
Female	89.600 (2.900)	93.666 (2.261)	0.750 (0.082)	0.821 (0.072)	0.571 (0.094)	0.786 (0.078)	28
F-test	[0.379]	[0.049]	[0.327]	[0.131]	[0.406]	[0.018]	
Panel B: Prize Category							
Chemistry	86.451 (1.329)	87.864 (1.331)	0.670 (0.035)	0.654 (0.033)	0.481 (0.037)	0.541 (0.036)	185
Physics	87.554 (1.245)	91.104 (1.247)	0.687 (0.032)	0.758 (0.031)	0.502 (0.034)	0.649 (0.034)	211
Medicine	87.585 (1.211)	88.897 (1.213)	0.691 (0.031)	0.726 (0.030)	0.534 (0.033)	0.650 (0.033)	223
Economics	86.038 (1.885)	88.156 (1.888)	0.587 (0.049)	0.696 (0.047)	0.413 (0.052)	0.511 (0.051)	92
F-test	[0.838]	[0.296]	[0.311]	[0.134]	[0.263]	[0.016]	
Panel C: Regions							
USA	83.829 (1.120)	86.597 (1.128)	0.584 (0.029)	0.639 (0.028)	0.396 (0.031)	0.525 (0.030)	255
Western Europe	88.743 (1.003)	90.570 (1.010)	0.701 (0.026)	0.736 (0.025)	0.531 (0.028)	0.632 (0.027)	318
Eastern Europe	92.801 (2.412)	94.753 (2.428)	0.818 (0.063)	0.891 (0.060)	0.727 (0.067)	0.727 (0.066)	55
Global South	87.739 (3.509)	88.398 (3.532)	0.769 (0.091)	0.846 (0.088)	0.538 (0.097)	0.692 (0.095)	26
All others	86.528 (2.370)	88.048 (2.385)	0.702 (0.062)	0.684 (0.059)	0.491 (0.065)	0.632 (0.064)	57
F-test	[0.002]	[0.013]	[0.002]	[0.001]	[0.000]	[0.016]	

Notes: The table shows group means for measures of Nobel laureate fathers' socioeconomic status in different subsamples. Standard errors are in parentheses. The first two columns show the average national income and education rank respectively, calculated as above. Columns 3 and 4 show the share of laureates in the given category who are classified in the top 10% of the income and education distributions, respectively. Columns 5 and 6 show the same for the top 5%. P-values from the F-test of equality of all group means are displayed below each set of means. Appendix Table A10 lists the countries in each region.

Table 2: Nobel Laureate Production by Commuting Zone, as a Function of Upward and Downward Mobility

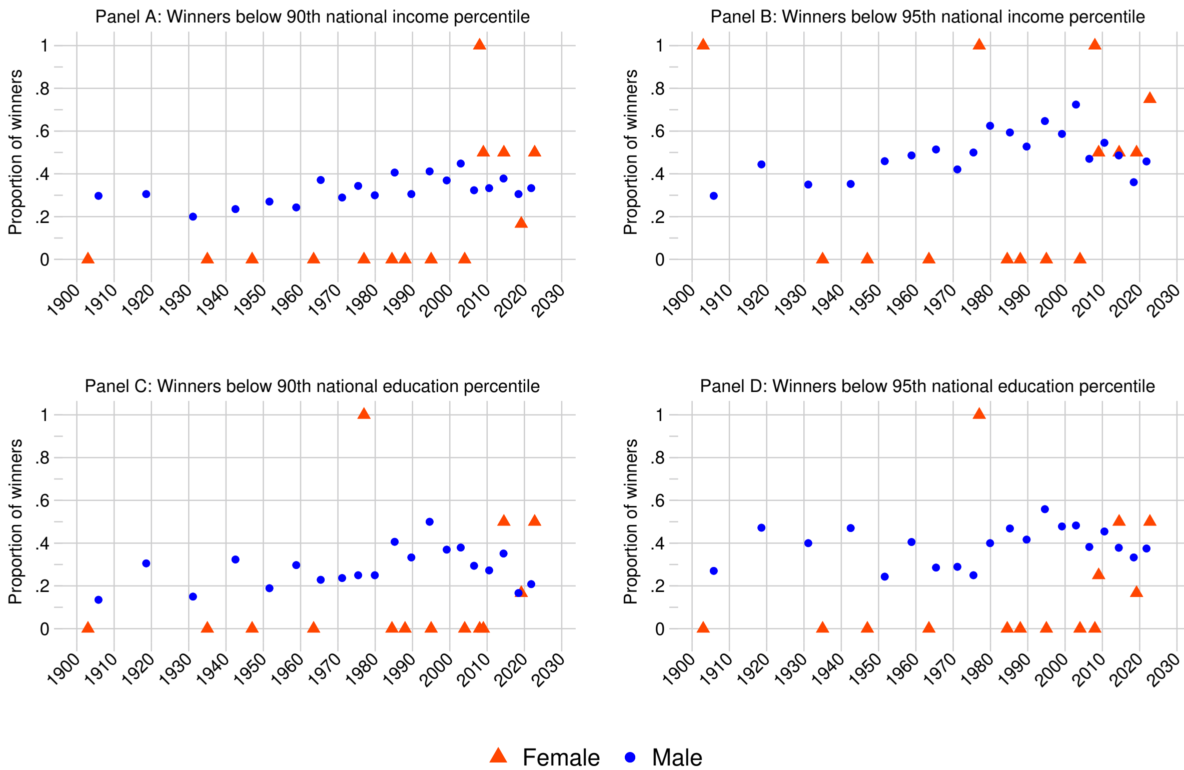
	Nr. Nobelists	Any winner	Below 90% Income	Below 90% Education	Below 95% Income	Below 95% Education
Expected rank of a child in	0.786	1.076***	0.763***	0.334*	0.801***	0.552***
a 25^{th} percentile family (p25)	(1.523)	(0.285)	(0.242)	(0.189)	(0.244)	(0.202)
Expected rank of a child in	-2.754	-0.992***	-0.650**	-0.229	-0.610**	-0.496**
a 75^{th} percentile family (p75)	(2.029)	(0.337)	(0.268)	(0.211)	(0.273)	(0.232)
Mean dependant variable	0.351	0.126	0.077	0.057	0.093	0.074
N (Commuting Zones)	740	740	740	740	740	740

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

Notes: The table reports estimates from an OLS regression of the number of Nobel laureates in the sciences emerging from a given commuting zone (CZ) in the United States, on two measures of intergenerational mobility in that CZ. Each observation is a commuting zone. p_{25} is absolute upward mobility, or the average income percentile for a child growing up in a 25th percentile family in the given CZ. p_{75} is an inverse measure of downward mobility, or the average income percentile for a child growing up in a 75th percentile family in the given CZ. Both are from the Opportunity Atlas (Chetty et al., 2018). The outcome in Column 1 is the number of Nobel laureates in the sciences in the CZ. In Column 2, the outcome is an indicator which takes the value 1 if a CZ has at least one winner. Similarly, the dependent variables in Columns 3–6 are indicators that take the value 1 if there is at least one winner from the given CZ with father income or education rank below either the 90th or 95th percentile. All specifications include quadratic population controls to account for the fact that Nobel Prize winners are more likely to come from bigger commuting zones.

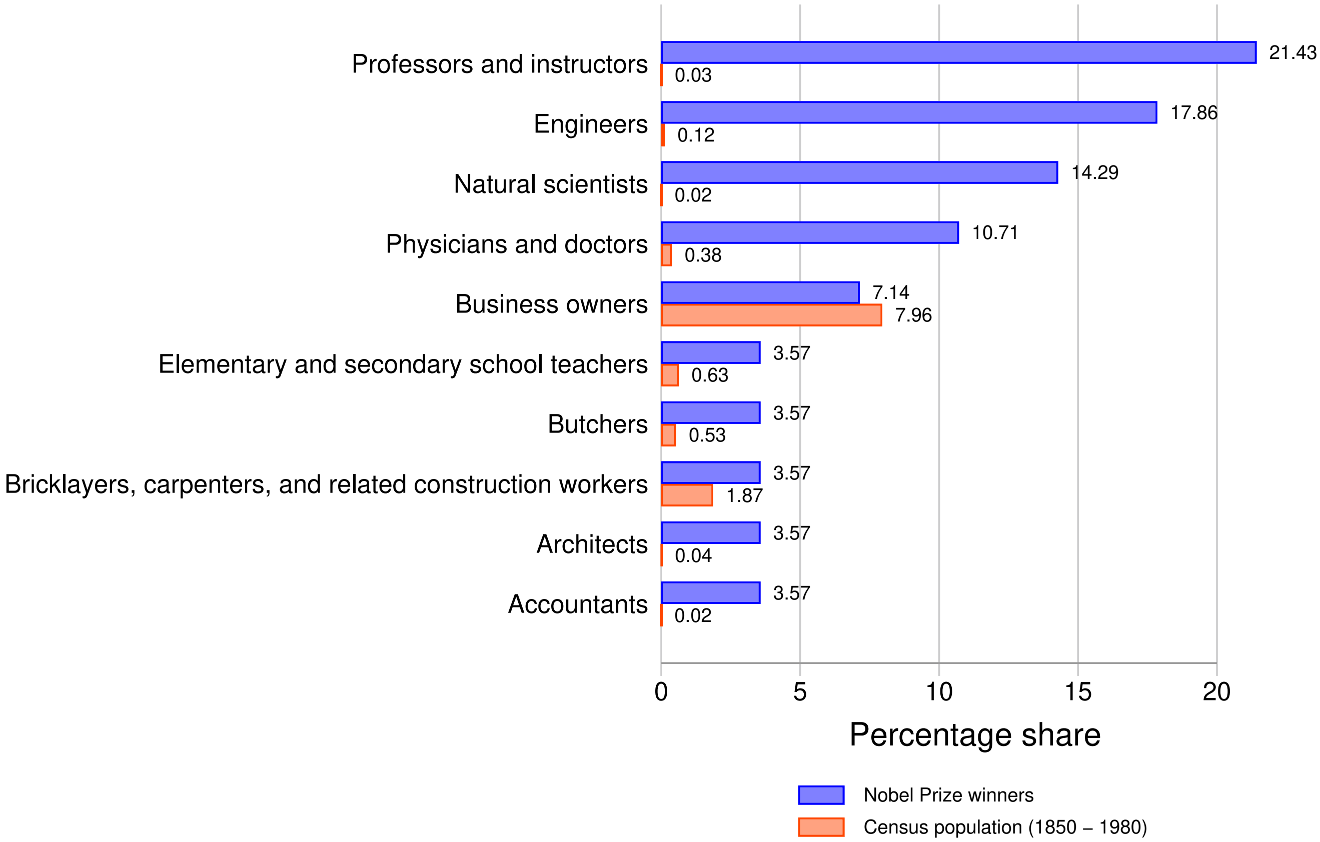
A Additional Figures and Tables

Figure A1: Share of Nobel Laureates with Fathers below a given SES Rank



Notes: The graph shows binned scatterplots of the proportion of winners in each time period with fathers below a given income and education percentile. Circles represent male laureates and triangles female laureates. Father income and education percentiles were predicted from father occupations, using data from the 1940 and 1950 U.S. Population Censuses.

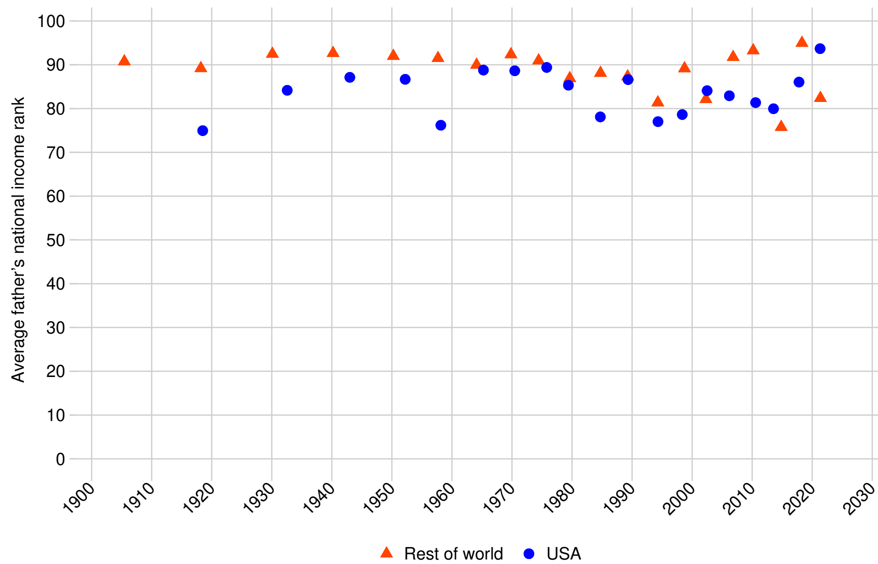
Figure A2: Most Frequent Father Occupations among Female Nobel Laureates



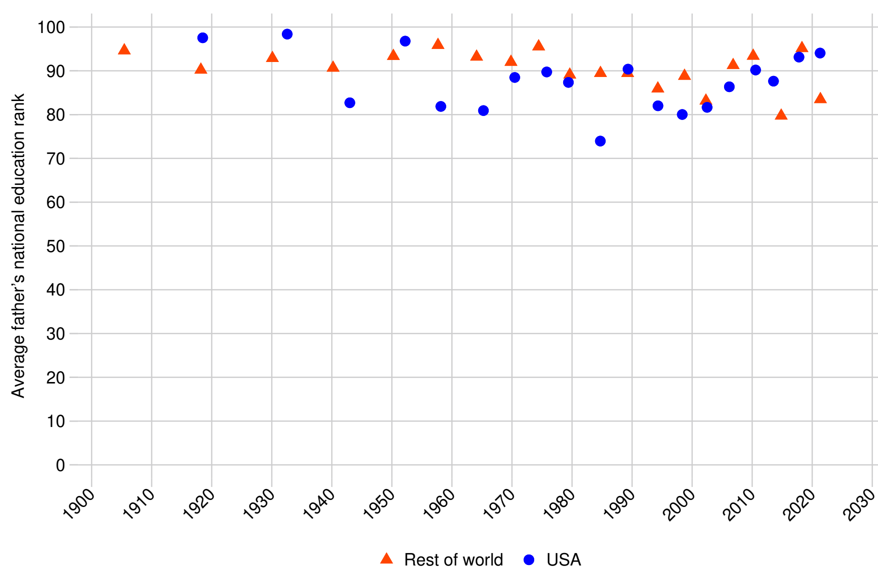
Notes: The graph shows the distribution of father occupations among *female* Nobel laureates in the sciences (in blue) and the general population (in orange). Occupational shares are averages from the U.S. Population Census between 1850 and 1980. The laureate sample covers the full set of prizes from 1901–2023.

Figure A3: Average SES Ranks of Fathers of Nobel Laureates: USA vs. Rest of World

A. Father income ranks



B. Father education ranks



Notes: The graph shows binned scatterplots of the proportion of winners in each time period with fathers below a given income (Panel A) and education (Panel B) percentile. Circles represent laureates who grew up in the United States, and triangles represent winners who grew up elsewhere. Father income and education percentiles were predicted from father occupations, using data from the 1940 and 1950 U.S. Population Censuses.

Table A1: Most Common Father Occupations of Nobel Laureates, by Prize Category

Father Occupation	Prize Category			
	Chemistry	Physics	Medicine	Economics
Business Owners	19%	14%	13%	15%
Physicians and doctors	9%	8%	15%	11%
Professors and academics	10%	14%	9%	5%
Engineers	8%	11%	8%	8%
Natural Scientists	3%	6%	4%	7%
Lawyers and Judges	4%	2%	5%	4%
Government Officials	5%	5%	7%	8%
Teachers	4%	7%	2%	5%
Retail Workers	3%	3%	3%	2%
Farmers	4%	2%	3%	4%
Observations	185	211	223	92

Notes: The table shows the ten most common occupations among fathers of the Nobel laureates in our sample, by prize category.

Table A2: Secular Trends in Average SES Ranks of Fathers of Nobel Laureates

	National Income Rank	National Education Rank
Constant (Year = 1901)	91.628***	93.715***
	(1.446)	(1.269)
Prize Year	-0.061***	-0.060***
	(0.019)	(0.018)
Observations	711	711

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

Notes: The table shows estimates from OLS regressions of father socioeconomic rank on prize year. The prize year variable has been rescaled to a start date of zero (representing 1901), so the constant can be interpreted as the fitted value for the first year of our sample. Standard errors are reported in parentheses. Father income and education percentiles were predicted from father occupations, using data from the 1940 and 1950 U.S. Population Censuses.

Table A3: Average SES Ranks of Fathers of Nobel Laureates and Time Trends: Robustness

	National Income Rank	National Education Rank
Panel A: Sample excluding difficult to classify observations		
Constant (Year = 1901)	91.743	94.154
	(1.466)	(1.222)
Prize Year	-0.061***	-0.064***
	(0.019)	(0.017)
Observations	698	698
Panel B: Sample excluding aristocrats		
Constant (Trend = 1901)	91.128	93.323
	(1.490)	(1.311)
Prize Year	-0.055***	-0.056***
	(0.019)	(0.018)
Observations	703	703
Panel C: Sample excluding business owners and farmers		
Constant (Trend = 1901)	92.345	95.735
	(1.371)	(1.119)
Prize Year	-0.052***	-0.049***
	(0.018)	(0.016)
Observations	580	580
Panel D: Sample excluding economics		
Constant (Trend = 1901)	91.712	93.943
	(1.473)	(1.290)
Prize Year	-0.062***	-0.064***
	(0.020)	(0.019)
Observations	619	619

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

Notes: The table shows estimates from an OLS regression of Nobel laureate parent SES rank on a linear time trend. These specifications are based on the primary specification in Appendix Table A2, with the following modifications. In Panel A, we exclude 13 laureates whose parent occupations were vaguely described and difficult to classify. In Panel B, we exclude 8 laureates whose parents were aristocrats (who were classified as being at education and income percentile 99 in the main analysis). In Panel C, we exclude 130 children of business owners and farmers, since these occupations can be associated with a wide range of possible SES ranks. In Panel D, we exclude the Nobel Prize in Economics.

Table A4: Average SES Ranks of Fathers of Nobel Laureates and Time Trends: Robustness to Alternate Codings

	National Income Rank	National Education Rank
Panel A: Heterogeneous coding of business owners		
Constant (Year = 1901)	91.464	93.281
	(1.461)	(1.318)
Prize Year	-0.063***	-0.065***
	(0.019)	(0.018)
Observations	711	711
Panel B: Recoding education according to biographies		
Constant (Trend = 1901)		93.454
		(1.291)
Prize Year		-0.062***
		(0.018)
Observations		711
Panel C: Recoding education and heterogeneous business owners		
Constant (Trend = 1901)		93.454
		(1.291)
Prize Year		-0.062***
		(0.018)
Observations		711

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

Notes: The table shows estimates from an OLS regression of Nobel laureate parent SES rank on a linear time trend. These specifications are based on the primary specification in Appendix Table A2, with the following modifications. In Panel A, we classified business owners as small, medium or large. For small business owners, we used the 25th percentile education and income data from the Census to calculate the father's rank. For large, we used the 75th. In Panel B, we override the father's education rank in cases where the historical record documents the actual education level of the father. In Panel C, we do both.

Table A5: Differences in Average SES Rank of Laureate Fathers, by Gender: Robustness

	Average Rank		**Top 10% Indicator**		**Top 5% Indicator**	
	National Income Rank	National Education Rank	National Income Rank	National Education Rank	National Income Rank	National Education Rank
Panel A: Sample excluding difficult to classify observations						
Constant (Year = 1901)	87.027	89.127	0.669	0.709	0.498	0.595
	(0.702)	(0.698)	(0.018)	(0.018)	(0.019)	(0.019)
Female Winners	3.985	4.573	0.109	0.105	0.095	0.220***
	(2.732)	(2.447)	(0.082)	(0.077)	(0.097)	(0.077)
Observations	698	698	698	698	698	698
Panel B: Sample excluding aristocrats						
Constant (Year = 1901)	86.823	88.874	0.664	0.705	0.486	0.591
	(0.701)	(0.708)	(0.018)	(0.018)	(0.019)	(0.019)
Female Winners	2.777	4.792	0.086	0.116	0.086	0.195**
	(2.984)	(2.370)	(0.084)	(0.075)	(0.096)	(0.080)
Observations	703	703	703	703	703	703
Panel C: Sample excluding business owners and farmers						
Constant (Year = 1901)	85.825	89.433	0.635	0.801	0.540	0.735
	(0.813)	(0.824)	(0.020)	(0.017)	(0.021)	(0.019)
Female Winners	3.732	5.326	0.134	0.083	0.076	0.111
	(3.227)	(2.435)	(0.085)	(0.065)	(0.098)	(0.073)
Observations	603	603	580	580	580	580
Panel D: Sample excluding economics						
Constant (Year = 1901)	86.823	88.874	0.677	0.710	0.502	0.608
	(0.747)	(0.753)	(0.019)	(0.019)	(0.021)	(0.020)
Female Winners	4.281	5.051	0.163**	0.130*	0.138	0.232***
	(2.936)	(2.532)	(0.076)	(0.076)	(0.098)	(0.076)
Observations	619	619	619	619	619	619

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

Notes: The table shows estimates from an OLS regression of Nobel laureate parent SES rank on an indicator for female laureates. The specifications are based on those in Panel A of Table 1, with the following modifications. In Panel A, we exclude 13 laureates whose parent occupations were vaguely described and difficult to classify. In Panel B, we exclude eight laureates whose parents were aristocrats (who were classified as being at education and income percentile 99 in the main analysis). In Panel C, we exclude 130 children of business owners and farmers, since these occupations can be associated with a wide range of possible SES ranks. In Panel D, we exclude the Nobel Prize in Economics.

Table A6: Differences in Average SES Rank of Laureate Fathers, by Gender: Robustness to Alternate Codings

	Average Rank		**Top 10% Indicator**		**Top 5% Indicator**	
	National Income Rank	National Education Rank	National Income Rank	National Education Rank	National Income Rank	National Education Rank
Panel A: Heterogeneous coding of business owners						
Constant (Year = 1901)	86.688	88.364	0.654	0.714	0.499	0.616
	(0.717)	(0.770)	(0.018)	(0.017)	(0.019)	(0.019)
Female Winners	2.815	2.706	0.060	0.107	0.072	0.169**
	(3.107)	(3.855)	(0.087)	(0.075)	(0.096)	(0.080)
Observations	698	698	711	711	711	711
Panel B: Recoding education according to biographies						
Constant (Year = 1901)		88.959		0.710		0.593
		(0.712)		(0.017)		(0.019)
Female Winners		2.631		0.076		0.193**
		(3.461)		(0.080)		(0.080)
Observations		698		711		711
Panel C: Recoding education and heterogeneous business owners						
Constant (Year = 1901)		88.959		0.710		0.593
		(0.712)		(0.017)		(0.019)
Female Winners		2.631		0.076		0.193**
		(3.461)		(0.080)		(0.080)
Observations		698		711		711

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

Notes: The table shows estimates from an OLS regression of Nobel laureate parent SES rank on an indicator for female laureates. The specifications are based on those in Panel A of Table 1, with the following modifications. In Panel A, we classified business owners as small, medium or large. For small business owners, we used the 25th percentile education and income data from the Census to calculate the father's rank. For large, we used the 75th. In Panel B, we override the father's education rank in cases where the historical record documents the actual education level of the father. In Panel C, we do both.

Table A7: Gender Differences in Laureate Father Ranks Before and After 2000

	Average Rank		**Top 10% Indicator**		**Top 5% Indicator**	
	National Income Rank	National Education Rank	National Income Rank	National Education Rank	National Income Rank	National Education Rank
Pre-2000	87.605	89.685	0.687	0.712	0.503	0.605
	(0.800)	(0.781)	(0.021)	(0.021)	(0.023)	(0.022)
Female Winners	9.719	7.334	0.313***	0.188*	0.297**	0.295***
	(1.299)	(1.707)	(0.021)	(0.097)	(0.129)	(0.098)
Post-2000	-2.210	-2.396	-0.069	-0.011	-0.039	-0.033
	(1.599)	(1.680)	(0.041)	(0.039)	(0.042)	(0.042)
Post-2000 x Female Winners	-9.805	-2.820	-0.320***	-0.112	-0.316*	-0.145
	(4.562)	(4.035)	(0.122)	(0.142)	(0.178)	(0.148)
Observations	711	711	711	711	711	711

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

Notes: The table reports estimates from OLS regressions of father SES rank on a female laureate indicator, interacted with an indicator for Nobel prizes awarded after 2000. The "female winner" shows that female winners consistently came from more elite families for prizes awarded before 2000. The point estimate of the interaction term is negative, suggesting that this effect may have diminished in recent years, but the term is statistically significant only in one out of six specifications. There are 10 female winners in our sample in the first 100 prize years (1901–1999), and 18 from 2000–2023.

Table A8: Secular Trends in Father SES Ranks of Nobel Laureates, by Prize Category

	National Income Rank	National Education Rank
Panel A: Chemistry		
Constant (Year = 1901)	93.952	94.917
	(2.179)	(1.955)
Prize Year	-0.103***	-0.097***
	(0.034)	(0.034)
Observations	185	185
Panel B: Economics		
Constant (Trend = 1901)	89.716	85.765
	(11.614)	(11.525)
Prize Year	-0.038	0.024
	(0.118)	(0.116)
Observations	92	92
Panel C: Medicine		
Constant (Trend = 1901)	93.720	94.832
	(2.721)	(2.488)
Prize Year	-0.086**	-0.084**
	(0.037)	(0.035)
Observations	223	223
Panel D: Physics		
Constant (Trend = 1901)	87.906	92.335
	(2.586)	(2.119)
Prize Year	-0.005	-0.017
	(0.034)	(0.030)
Observations	211	211

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

Notes: The table shows OLS regression estimates of laureate father SES rank on prize year, separately for each prize category. The prize year variable has been rescaled to a start date of zero, so the constant is interpreted as the fitted value for the father ranks of laureates in 1901. Standard errors are in parentheses.

Table A9: Father Ranks of Nobel Laureates: USA vs. Rest of World

	Average Rank		**Top 10% Indicator**		**Top 5% Indicator**	
	National Income Rank	National Education Rank	National Income Rank	National Education Rank	National Income Rank	National Education Rank
Constant	88.899	90.636	0.719	0.754	0.550	0.647
	(0.792)	(0.762)	(0.021)	(0.020)	(0.023)	(0.022)
US-born Winners	-5.069	-4.038	-0.135***	-0.115***	-0.154***	-0.121***
	(1.458)	(1.509)	(0.037)	(0.036)	(0.039)	(0.039)
Observations	711	711	711	711	711	711

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

Notes: The table shows estimates from OLS regressions of SES ranks of fathers of Nobel laureates on an indicator that takes the value one for laureates who spent their childhood in the United States. Standard errors are reported in parentheses.

Table A10: Countries in each region

West	East	Global South
Argentina	Azerbaijan	Algeria
Australia	Belarus	Brazil
Austria	Bosnia and Herzegovina	Egypt
Belgium	China	India
Canada	Croatia	Indonesia
Cyprus	Czech Republic	Lebanon
Denmark	Hungary	Mexico
Finland	Latvia	Morocco
France	Lithuania	Pakistan
Germany	Poland	South Africa
Ireland	Romania	South Korea
Israel	Russia	Taiwan
Italy	Slovakia	Turkey
Japan	Slovenia	Venezuela
Luxembourg	Ukraine	
Netherlands		
New Zealand		
Norway		
Portugal		
Spain		
Sweden		
Switzerland		
United Kingdom		
United States of America		

Notes: The table lists the countries included in each of the geographical regions titled "Western Europe", "Eastern Europe", "Global South", and "All others" that are used in the analysis for Table 1.

Table A11: Laureate Production and Commuting Zone Intergenerational Mobility: Alternate Mobility Measures

	Nr. Nobelists	Any winner	Below 90% Income	Below 90% Education	Below 95% Income	Below 95% Education
% Δ earnings low-income children	0.604	0.975***	0.654***	0.391***	0.742***	0.568***
	(0.746)	(0.204)	(0.178)	(0.143)	(0.184)	(0.159)
% Δ earning high-income children	-4.365**	-1.732***	-1.125***	-0.595**	-1.214***	-0.949***
	(1.729)	(0.393)	(0.340)	(0.257)	(0.350)	(0.297)
Mean dependant variable	0.351	0.126	0.077	0.057	0.093	0.074
N (Commuting Zones)	714	714	714	714	714	714

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

Notes: The table shows estimates analogous to those in Table 2, but using alternate measures of p_{25} and p_{75}. The estimates in Table 2 use descriptive measures of p_{25} and p_{75}, which describe average outcomes of children born in each commuting zone at the first and third quartiles. The estimates in this table instead use the "causal" measures of p_{25} and p_{75}, which are estimates of the marginal effect on childhood SES rank of spending twelve additional years in the given CZ, estimated from individuals who moved between commuting zones as children. We use the descriptive estimates in the primary analysis as they are more precise, being drawn from larger samples of individuals, but the estimates here are largely similar to those in Table 2. All specifications include quadratic population controls to account for the fact that Nobel Prize winners are more likely to come from bigger commuting zones. Standard errors are in parentheses.

Table A12: Laureate Production and Local Intergenerational Mobility: Descriptive County Estimates

	Nr. Nobelists	Any winner	Below 90% Income	Below 90% Education	Below 95% Income	Below 95% Education
Expected rank of a child in	0.268*	0.169***	0.165***	0.076**	0.137***	0.097**
a 25^{th} percentile family (p25)	(0.141)	(0.051)	(0.045)	(0.038)	(0.045)	(0.041)
Expected rank of a child in	-0.667***	-0.039	-0.079*	-0.039	-0.021	-0.044
a 75^{th} percentile family (p75)	(0.225)	(0.053)	(0.047)	(0.037)	(0.047)	(0.043)
Mean dependant variable	0.079	0.036	0.021	0.016	0.026	0.021
N (Counties)	3121	3121	3121	3121	3121	3121

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

Notes: The table presents results from the same specifications as Table 2 but using county-level laureate production and mobility measures instead of commuting zone. All specifications include quadratic population controls to account for the fact that Nobel Prize winners are more likely to come from bigger counties. Standard errors are in parentheses. The county measures of both mobility and laureate production are less precise, notably because over 95% of counties produce zero laureates, vs. 87% of CZs.

Table A13: Laureate Production and Local Intergenerational Mobility: Causal County Estimates

	Nr. Nobelists	Any winner	Below 90% Income	Below 90% Education	Below 95% Income	Below 95% Education
% Δ earnings low-income children	0.432***	0.265***	0.201***	0.120***	0.189***	0.141***
	(0.153)	(0.056)	(0.052)	(0.045)	(0.052)	(0.049)
% Δ earning high-income children	-2.037***	-0.195	-0.223	-0.170	-0.104	-0.180
	(0.581)	(0.150)	(0.137)	(0.113)	(0.138)	(0.128)
Mean dependant variable	0.079	0.036	0.021	0.016	0.026	0.021
N (Counties)	2869	2869	2869	2869	2869	2869

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

Notes: The table presents results from specifications identical to those in Table A12, but using the causal mobility measures from Chetty and Hendren (2018b). See the note on Appendix Table A11 for more details on the difference between the causal and descriptive mobility estimates. All specifications include quadratic population controls to account for the fact that Nobel Prize winners are more likely to come from bigger counties. Standard errors are in parentheses.

B Additional Information on Data and Methods

B.1 Biographical Data

We hand-collected biographical information for every Nobel laureate in Physics, Chemistry, Physiology, and Economics. For each laureate, we reviewed multiple sources to identify their birth dates, birth locations, and the occupations of their parents. The availability of parent occupational information was heterogeneous. Many winners described their parents work in their acceptance speeches or biographies. Parent occupations were also recorded in documents like obituaries, biographies, scientific directories and encyclopedia entries. When we could not find information on father occupation, we reached out to living laureates, a handful of whom responded with information about their parents. We obtained data on father occupations for 715 out of 739 laureates (96.7%), but only 181 mother occupations (25.3%). Mothers' occupations are also generally less informative about laureates' childhood socioeconomic status, because so many women in this era were out of the labor force.

When laureates moved before the age of six, we assigned them to their location at age six. Results were unchanged if we used other ages as anchors for their home country. When laureates were born to expatriates (i.e. parents living temporarily in countries other than their home country), we assigned them to their parents' home countries, as the home country occupational distributions are likely to be more relevant for determining parental rank. We similarly assigned them to home countries when parents were serving in a foreign service. For example, Ronald Ross (Physiology, 1902) was the son of a British general serving in India; at age 8, he moved to the Isle of Wight for school. Our assumption is that his parents' socioeconomic status is better characterized by the average status of an army general in the United Kingdom than that of an army general in India. There are only a handful of cases like these, so they do not have substantial bearings on our results.

We construct cohorts of winners by rounding each winner's birth year to the nearest decade (e.g., a winner born in 1893 belongs to the 1890 cohort, whereas a winner born in 1898 belongs to the 1900 cohort). There are six winners born between 1835 - 1843, all of whom are initially assigned to the 1840 cohort. However, due to the lack of availability of census data during these years, we then map these observations to 1850 instead, the closest available decade.

B.2 Occupational Earnings and Education from U.S. Population Census

To estimate the socioeconomic status of laureates' parents, we aimed to identify the average income and education ranks of people with the same occupation as each laureate's father in the U.S. Census. We used data from the 1850–1970 U.S. population censuses, extracted from IPUMS and proceeded as follows.

We restricted the census sample to working age (18–65 years old) men, and dropped all individuals who were not in the labor force. We also dropped census observations for which the occupational category variable takes on the following values: "students", "disabled individuals with no reported occupation", "other non-occupation", "n/a (blank)". Excluding the unemployed could marginally bias our ranks downward, but the bias should be small, given that (i) the historic U.S. unemployment rate has typically fluctuated between only 5–8%; and (ii) the unemployed are distributed across many different occupations, limiting the degree of bias.

We matched each laureate's father occupation to the nearest harmonized occupational category (*occ1950*, created by IPUMS). When fathers held multiple occupations, we assigned the one described as the primary occupation during the laureate's childhood years. Appendix Table A14 shows the list of occupational categories used in the sample.

There is inevitably some measurement error in classifying business owners, as the historical record was often unclear on the scale of the business; as such, a business owner could represent a relatively upper or middle class family. We classified business owners with code 290, which is "Managers, officials, and proprietors, not otherwise classified." This puts business owners, on average, at the 86th education percentile and the 93rd income percentile. Farm owners exist as an occ1950 category, but similarly cover a wide range of socioeconomic ranks. To ensure that our assumptions about these occupational ranks were not driving any of our results, we verified that our primary results are similar when we exclude business owners and farmers (Appendix Tables A3 and

A5).

Individual and household earnings were available beginning in 1950, and years of education in 1940, so we used these values for all earlier years. We excluded negative or zero earnings values. Literacy is recorded in earlier census years, but is considered less reliable by historians, as collection methodology varied substantially across enumerators (Collins and Margo, 2006). For decades where these observations were available, we applied a three decade moving average to smooth out fluctuation in occupational outcomes from sampling variation. We similarly interpolated missing values (e.g. where there were no observations in a given occupation-decade cell). These adjustments have little bearing on our results, as they only affect census years after 1950 (for education) and 1960 (for earnings), and the vast majority of Nobel laureates were born in earlier decades.

We also estimated the population share of each occupation in each decade, which is essential for calculating ranks. To calculate ranks, we assigned individuals an earnings level based on their occupation, and then ordered individuals within each decadal census according to those earnings levels. We then assigned the mean earnings rank within each occupation group, and collapsed the data to occupation-decade level. These earnings ranks were then matched to the Nobel laureates database at the occupation-decade level. We performed a similar exercise for education ranks.

Since occupational incomes were not observed before 1950, there is no change in the relative ordering of occupations before 1950. However, we observe the share of people with each occupation in all census years, so the *ranks* of occupations do change continuously over time. For instance, trade workers tended to have relatively high education ranks in 1850, but had lower ranks by 1950, due to the growing number of more educated white collar workers.

B.3 Historical Estimates of Country-Level GDP per Capita and Population

We used country of childhood in two ways in our analysis. For our primary analysis, we used country of childhood to adjust the relative occupational distributions, to reflect the fact that an individual with only some schooling (say, literacy) would occupy a higher education rank in a poor country than in a rich country. Second, for the section on income differences across countries only, we used national GDP per capita to adjust occupational earnings for living standards across countries. This second adjustment reflects the fact that an individual at the 90th percentile of the earnings distribution in India has a much lower global income rank than an individual at the 90th percentile in the United States.

We used historical estimates of GDP per capita and population for 217 countries from 1830–1970, from Fariss et al. (2022), with adjustments described below.

B.3.1 Adjusting for Occupation Distributions

The distribution of occupational population shares affects the resulting socioeconomic rank of each occupation, even if the ordering of occupations is unchanged. For example, the percentile rank of a sawyer in the U.S. is much higher in 1850, a time when roughly 44% of the population were farmers, compared to 1950, when that share is down to only 9%. Similarly, countries will have different occupational distributions at different levels of economic development. However, we only had data on the historical occupational distribution of the United States.

We therefore used the U.S. historical occupational distribution to proxy for the occupational distribution in other countries at different points in time. We assigned each country the U.S. occupational distribution from the decade when its development level (measured by GDP per capita) was nearest to that of the U.S., using minimum absolute distance.

Consider an example of how we would calculate the earnings rank of an Indian tailor in 1950. First, we identified the U.S. decade in our sample data with the nearest GDP per capita level to that of India in 1950 — this was 1850. In 1850 in the U.S., a tailor had an earnings rank of 84; we assign this rank to the Indian tailor for the primary analysis. By contrast, in the U.S. in 1950, a tailor would have an earnings rank of only 55.

B.3.2 Calculating Global Income Ranks

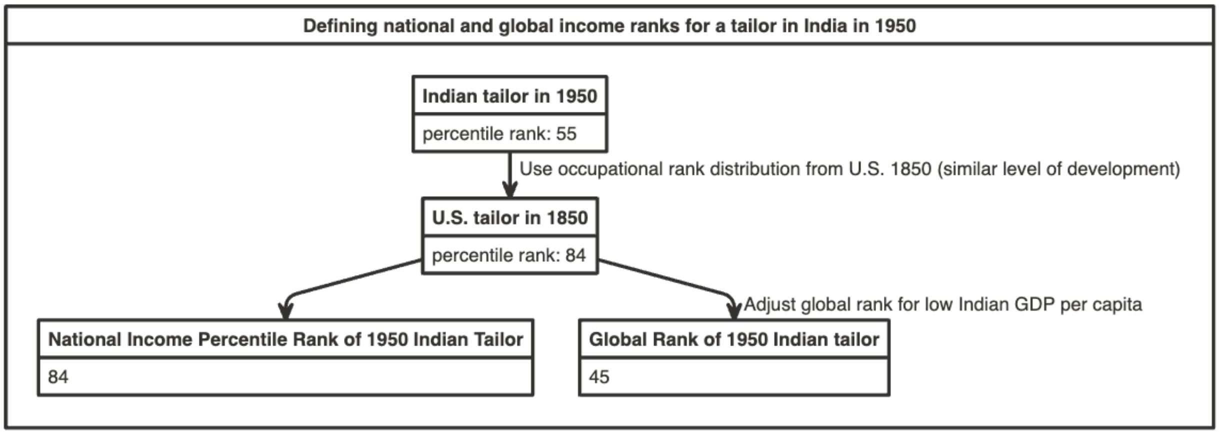
Our primary analysis focuses on occupational ranks within countries. Since most winners are from the West, this tells us how effectively the West is at mobilizing its own talent. In the final part of the results, we instead consider global income ranks, which tell us how effective human society is at mobilizing global talent. Consider again the Indian tailor in 1950 above. The tailor occupied the 84th earnings rank in India, but their global earnings rank was much lower, since India in 1950 was a very poor country.

We calculate a global earnings rank as follows. First, for each country-decade-occupation group, we scale the occupational income by the GDP per capita ratio of each country with the United States in the same decade. This gives us a real income estimate for each country-decade-occupation group. Second, we create a synthetic global population based on each country's occupation shares (adjusted for development level, as above) and their population. An individual's rank in this synthetic population is that individual's global income rank.

Let us continue with the example of the Indian tailor above. We have already estimated that the tailor is at the 84th earnings percentile in India in 1950. In 1950, U.S. GDP per capita was approximately 16 times higher than Indian GDP; we therefore multiply the Indian tailor's predicted income by 0.06. After conducting this calculation for every country-occupation-decade, the Indian tailor has a global income rank of 45. This reflects the fact that the tailor is relatively well off in India, but India is relatively poor globally. Appendix Figure A4 summarizes the example.

For individuals from poor countries, global ranks are systematically lower than domestic ranks, and vice versa. A U.S. tailor in 1950 would be at the 97th percentile globally. Since most of the Nobel laureates are from rich countries, the average global earnings rank is therefore higher than the average domestic earnings rank.

Figure A4: Example Calculation of National and Global Earnings Ranks



B.4 Opportunity Atlas Data

U.S. born winners account for 35% of the total sample. To carry out additional analyses on this subsample we supplemented our data with two measures of access to opportunity from the Opportunity Atlas (Chetty et al., 2018). The Opportunity Atlas describes upward mobility (p_{25}) as the expected earnings rank of a child born to a household at the 25th percentile of the earnings distribution. These are calculated for U.S. children born in 1980 from the universe IRS tax records. The Atlas also reports an inverse measure of downward mobility (p_{75}), which is the expected earnings rank of a child born to a household at the 75th percentile of the earnings distribution. Note that if (p_{25}) is low, children born poor are likely to stay poor, while if (p_{75}) is low, then children born rich are *unlikely* to stay rich. We focus our description here on (p_{25}); the data construction for (p_{75}) is analogous.

We match U.S. laureates to childhood commuting zones (CZs) and counties based on the biographical data. The Opportunity Atlas reports two different measures of (p_{25}) for each CZ and county. The descriptive measure is simply the average earnings rank of children born at the 25th percentile in a given CZ. The 10th percentile CZ has $p_{25} = 0.36$, while the 90th has $p_{25} = 0.52$.

This is a precise measure, due to the large number of observations in each CZ in the tax data, and it is our primary measure. The limitation of the descriptive measure is that it does not necessarily describe the *causal effect* of growing up in a given CZ, because upwardly mobile parents may select into CZs with other upwardly mobile people.

The second Opportunity Atlas measure is the *causal* upward mobility measure. To calculate causal effects of CZs on upward mobility, Chetty and Hendren (2018a) examine families who moved across different CZs. In these families, younger children had more exposure to the new CZ than older children. From these data, Chetty and Hendren (2018a) calculate the *exposure effect* of each CZ, the effect on adult income rank of spending one more year of childhood in a given CZ. To make these exposure effects comparable in magnitude to the descriptive estimates above, we multiply them by 12, giving us a measure of 12 years of exposure in the laureate's childhood CZ (representing approximately 12 years of pre-college education). After this rescaling, The 10th percentile CZ has $p_{25} = -0.06$ (6 rank points worse than the mean CZ), while the 90th has $p_{25} = 0.14$. The causal and descriptive measures are highly correlated across CZs ($\rho = 0.9$).

While the latter measure is more credibly causal, it is less precise for each county and CZ as it is identified on a much smaller set of individuals. Our primary analysis (Table 2) uses the descriptive CZ estimates. We show robustness to using the causal estimates (Appendix Table A11) and to using both sets of estimates at the county level (Appendix Tables A12 and A13).

Table A14: List of occupational categories present in the analysis sample

Occupation Title	Frequency	Percent
Commercial Managers	105	15.09
Health professionals	81	11.64
Professors and instructors	69	9.91
Engineers	60	8.62
Natural scientists	34	4.89
Elementary and secondary school teachers	32	4.60
Jurists	28	4.02
Officials (government and non-profit organizations)	28	4.02
Farmers and farm managers	23	3.30
Sales workers and shop assistants	22	3.16
Accountants	20	2.87
Workers in religion	17	2.44
Bricklayers, carpenters, and related construction workers	14	2.01
Insurance agents	10	1.44
Other agents	9	1.29
Journalists, authors, and related writers	8	1.15
Protective service workers	8	1.15
Creative artists	7	1.01
Statistical and social scientists	7	1.01
Office and clerical workers	6	0.86
Operatives and kindred workers	6	0.86
Professional, technical, and related workers	6	0.86
Tailors and related workers	6	0.86
Welders and related metal workers	6	0.86
Architects	5	0.72
Forestry workers	5	0.72
Members of the armed forces	5	0.72
Bakers	4	0.57
Blacksmiths and machinists	4	0.57
Building managers and proprietors	4	0.57
Craftsmen and kindred workers	4	0.57
Jewelers, opticians, and precious metal workers	4	0.57
Locomotive operators	4	0.57
Other mechanics	3	0.43
Plumbers and pipe-fitters	3	0.43
Real estate agents	3	0.43
Ship officers	3	0.43
Social and welfare workers	3	0.43
Telephone operators	3	0.43
Textile workers	3	0.43
Vehicle mechanics	3	0.43
Aircraft pilots and navigators	2	0.29
Bookkeepers and related workers	2	0.29
Printers and related workers	2	0.29
Truck drivers	2	0.29
Butchers	1	0.14
Electronic service and repair workers	1	0.14
Farm laborers	1	0.14
Food service workers	1	0.14
Foremen	1	0.14
Gardeners	1	0.14
Launderers and dry-cleaners	1	0.14
Miners and related workers	1	0.14
Newsboys and deliverymen	1	0.14
Nonmedical technicians	1	0.14
Painters	1	0.14
Sawyers and lumber inspectors	1	0.14
Transport conductors	1	0.14

Notes: The table lists the occupational categories (according to the 1950 census classification) in our analysis sample and their frequency among the Nobel winners' fathers.