NBER WORKING PAPER SERIES

INEQUALITY IN SCIENCE:
WHO BECOMES A STAR?

Anna Airoldi
Petra Moser

Working Paper 33063
http://www.nber.org/papers/w33063

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2024

Moser gratefully acknowledges financial support from the National Science Foundation through award # 1824354 “Social Mobility and the Origins of American Science” and from NYU’s Center for Global Economy and Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Inequality in Science: Who Becomes a Star?
Anna Airoldi and Petra Moser
NBER Working Paper No. 33063
October 2024
JEL No. J24, N0, N32, O3

ABSTRACT

How does a person's childhood socioeconomic status (SES) influence their chances to participate and succeed in science? To investigate this question, we use machine-learning methods to link scientists in a comprehensive biographical dictionary, the American Men of Science (1921), with their childhood home in the US Census and with publications. First, we show that children from low-SES homes were already severely underrepresented in the early 1900s. Second, we find that SES influences peer recognition, even conditional on participation: Scientists from high-SES families have 38% higher odds of becoming stars, controlling for age, publications, and disciplines. Using live-in servants as an alternative measure for SES confirms the strong link between childhood SES and becoming a star. Applying text analysis to assign scientists to disciplines, we find that mathematics is the only discipline in which SES influences stardom through the number and the quality of a scientist's publications. Using detailed data on job titles to distinguish academic from industry scientists, we find that industry scientists have lower odds of being stars. Controlling for industry employment further strengthens the link between childhood SES and stardom. Elite undergraduate degrees explain more of the correlation between SES and stardom than any other control. At the same time, controls for birth order, family size, foreign-born parents, maternal education, patents, and connections with existing stars leave estimates unchanged, highlighting the importance of SES.

Anna Airoldi
Revelio Labs
New York, NY 10012
aa6330@nyu.edu

Petra Moser
Department of Economics
NYU Stern
44 West 4th Street
New York, NY 10012
and NBER
pmoser@stern.nyu.edu

A data appendix is available at http://www.nber.org/data-appendix/w33063

As talented women and Black men have entered high skilled professions, improvements in the allocation of talent have contributed 20-40% of productivity growth since the 1960s (Hsieh et al. 2019). Yet, children from low-income families continue to be underrepresented among inventors (Bell et al. 2019, Aghion et al. 2018), PhD recipients (Stansbury and Schultz 2023), professors (Morgan et al. 2022), and in creative occupations (Biasi, Deming, and Moser 2022; Biasi, Dahl and Moser 2024).

This paper investigates the influence of socioeconomic status (SES) in science, specifically, whether a person's childhood SES influences their chances of becoming a "star."
While stardom is an important factor in hiring and promotion, it is difficult to quantify because conversations about stardom typically happen behind closed doors and are rarely recorded. We address this data challenge by exploiting a unique feature of *American Men of Science* (MoS 1921), an exceptionally comprehensive biographical dictionary of "all persons in the nation who were contributing to science" (Rossiter 1982, p. 25). To collect data for his own analyses of intelligence, the original editor of the MoS, James McKeen Cattell asked scientists to rank each other. Based on these rankings, Cattell placed a star (*) next to the entries of roughly 1,000 "leading men of science." We hand-collect these data, along with each scientist's full name, exact birth date and birthplace, university education, employment, and research topics, and use machine-learning algorithms to match scientists with the census record of their childhood home. Owing to the rich biographical information in the MoS, we achieve exceptionally high (84%) linking rates; 15.5% of the linked scientists are stars.

Examining selection into the field of science, we show that patterns of underrepresentation that affect innovation today (e.g., Bell et al. 2019; Aghion et al. 2018) already existed more than a century ago. Children of attorneys, physicians, and clergymen are overrepresented in science, while children of farmers, farm laborers, and other low-SES jobs are severely underrepresented. Using fathers' occupational income ranks (OCCSCORE) to measure childhood SES, we show that the median scientist is drawn from the 78th percentile of the income distribution (equivalent to the rank of a salesperson or carpenter), while the median boy of the same age in the population is the son of a farmer. Alternative measures of SES that capture differences in education and occupational prestige imply even greater underrepresentation of people with low childhood SES.[1]

To complement composite measures, such as the OCCSCORE, we create a simple, direct measure of SES by recording the presence of live-in servants in a person's childhood home. This alternative measure confirms severe underrepresentation by SES: Scientists are nearly 20 times as likely to have grown up in households with servants compared with boys of the same age in the population. 8.6% of scientists whom we observe as children in the census of 1880 grew up in households with servants, compared with just 0.4% of same-aged boys in the population. 20 years later, in 1900, 24.1% of scientists lived in households with servants, compared with just 1.4% in the population.

Extending our analyses, we investigate whether SES matters beyond participation, by influencing peer recognition in science. Research in social psychology suggests that inequality influences how we perceive wealthy and poor individuals (e.g., Jetten et al. 2017; Moya and Fiske 2017) and, specifically, how highly we rank them on competence and assertiveness, traits that allow individuals to "get ahead" (Tanjitpiyanond, Jetten, and Peters 2022).[2] If scientists from poor families are less likely to be considered competent and assertive, they may also be less likely to be leaders in their fields, or "stars." We use linked data on scientists, their education, jobs, publications, and childhood SES to investigate this hypothesis.

Consistent with a link between childhood SES and peer recognition, we find that scientists from high-SES families are consistently more likely to be stars. Using the occupational income score (OCCSCORE) of the scientist's father as a measure of SES, we find that children from the top half of the occupational income distribution have 38.7% higher odds of becoming stars. Similarly, we find that scientists from households with servants have 75.7% higher odds of becoming stars. Alternative measures of childhood SES confirm these findings, with 35.0% to 53.0% higher odds of stardom for scientists from high-SES homes.

---

[1] See the "Integrated Occupation and Industry Codes and Occupational Standing Variables in the IPUMs", as well as our data section for a discussion of the OCCSCORE and alternative measures of SES.

[2] This literature classifies stereotypes in a vertical and horizontal dimension: The vertical dimension covers a group's competence and assertiveness – that is the ability of its group members to "get ahead" and achieve higher status in society (Abele, Ellemers, Fiske, Koch, and Yzerbyt 2021), while the horizontal dimension captures a group's friendliness and morality, or the prosocial tendencies of group members to "get along." Evidence on the impact of inequality on stereotyping is stronger on the vertical dimension (Heiserman and Simpson 2017).

To extend these findings, we investigate alternative mechanisms by which SES may influence peer recognition in science. For example, scientists from high-SES families may become stars because high SES is associated with better health, cognitive, and socioemotional outcomes (e.g., see Bradley and Corwyn 2002). Similarly, high-SES scientists may attend better universities (Chetty et al. 2017), which are closer to the knowledge frontier (Biasi and Ma 2023), preparing them to write more and better papers.

The link between SES and stardom, however, is robust to controlling for differences in the number and the quality of publications. Controlling for publications and citations, scientists from high-SES families have 38.3% higher odds of becoming a star (just slightly below the 38.7% without controls for publications), and scientists from households with servants have 55.8% higher odds (compared with 75.7%). Moreover, all results are robust to controlling for alternative measures for the quality of publications.

Importantly, the influence of childhood SES may vary across disciplines depending on variation in the intensity of classism or in the importance of early childhood investments (e.g., National Research Council 2009). Social psychologists have found that social class stereotyping is stronger in societies with greater degrees of inequality (e.g., Tanjitpiyanond et al. 2022), and disciplines with greater degrees of inequality may similarly stereotype more, creating a "leg up" for high-SES scientists in becoming stars. To investigate differences across disciplines, we use natural language processing to assign scientists uniquely to 12 disciplines in which Cattell identified stars: anatomy, anthropology, astronomy, botany, chemistry, geology, mathematics, pathology, physics, physiology, psychology, and zoology.

Re-estimating the correlation between SES and stardom within disciplines, we find that for mathematics – but not other disciplines – SES influences stardom through publications. Without controls for publications, high-SES mathematicians have 73.5% higher odds of becoming stars. Interestingly, that link disappears with controls for publications, suggesting that SES may influence peer recognition through publications. This result, however, is unique to math; in other disciplines, the correlation between SES and recognition is robust to controlling for publications. In pathology, for example, scientists from high-SES families have nearly three times higher odds of becoming stars, controlling for publications.

In addition to disciplines, the types of jobs that scientists hold may influence whether they are perceived as stars, irrespective of publications. Specifically, scientists whose advisors have pushed them to pursue academic jobs may attach more prestige to university employment. In contrast, SES may play a larger role in industry if jobs are passed on through family networks (e.g., San 2023) or if research productivity in industry is more difficult to observe, leaving more room for bias. To control for such differences, we use career titles to distinguish academic from industry scientists. This analysis reveals that industry scientists have 66.6% lower odds of stardom. Moreover, controlling for industry employment strengthens the link between childhood SES and stardom: high-SES scientists have 46.9% higher odds of becoming stars controlling for industry jobs and publications, and scientists from households with servants have 66.7% higher odds. Industry scientists are also more likely to come from high-SES families, and the link between childhood SES and stardom is stronger in industry. Taken together, these results suggest that despite inequities, academia is more egalitarian than industry.

Elite undergraduate degrees explain more of the correlation between SES and stardom than any other controls. Yet, even controlling for elite undergraduate degrees, scientists from high-SES families have 43.3% higher odds of being stars, and scientists with servants have 58.1% higher odds controlling for publications and industry employment. Controlling for elite graduate degrees leaves the estimate at 44.9% higher odds for scientists from high-OCCSCORE homes and 66.6% for those from households with servants. By comparison, controls for patents, connections with existing stars, birth order, family size, parents' immigration status, and mothers' education leave estimates substantially unchanged.

Farmers account for a disproportionate share of the population and scientists in our data. In the 1870 census, 31.1% of scientists and 45.6% of boys of the same age in the population are the sons of farmers. Did SES, and more specifically, personal wealth, influence the odds that the son of a farmer would become a star? We investigate this question using data on personal wealth in the census of 1870, and again confirm the importance of SES: The sons of farmers in the top 2.5% of personal wealth have 60.5% higher odds of becoming stars.

Taken together, our findings indicate that patterns of inequality that affect science and innovation today already existed in the early 20th century, and that inequality extends to peer recognition, even conditional on participation. Decomposing the sources of U.S. productivity growth, Hsieh et al. (2019) show that 20 to 40% of productivity growth between 1960 and 2010 is due to improvements in the allocation of talent, as talented women and Black men have entered high-skilled professions. Yet, linking tax records with patents for children born 1980-84, Bell et al. (2019) show that children in the top 1% of parental income are 10 times more likely to become inventors than children in the bottom half. Examining inventors in Finland, Aghion et al. (2018) find that controlling for parental SES, education, and children's IQ weakens the correlation between parental income and patenting. Analyzing linked data on innovators in the United States and Finland, Einiö, Feng, and Jaravel (2024) show that innovators create products for consumers like them in terms of SES, gender, and age. Our findings complement analyses of contemporary data by investigating the historical roots of socioeconomic inequality and by documenting inequality in peer recognition.

Existing research on star scientists has analyzed the positive spillovers that stars create for their collaborators (e.g., Azoulay, Graff Zivin, and Wang 2010; Akcigit et al. 2018) and investigated the effects of taxation on the location decisions of star inventors. Azoulay et al. (2010) exploit the unexpected death of 112 academic superstars to investigate changes in the productivity of coauthors after the death of a star and document a lasting 5 to 8% decline in quality-adjusted publications for coauthors. Examining teams of inventors in European patent applications, Akcigit et al. (2018) show that inventors produce more patents after they interact with star inventors. Akcigit, Baslandze, and Stantcheva (2016) use U.S. and European patents to investigate the effects of top tax rates in the international mobility of inventors and show that inventors' location decisions are significantly affected by top tax rates. Investigating the effect of taxes on the location decisions of star inventors across U.S. states, Moretti and Wilson (2017) show that star inventors move from states with high tax rates to lower tax locations. In this literature, star scientists are typically defined by exceptional productivity, e.g., in terms of patents and publications. To these findings, we add an analysis of peer recognition, conditional on productivity.

Examining peer recognition in economics, Card et al (2022) show that women with comparable publications and citations were less likely to be elected Fellows of the Econometric Society between 1933 and 1980. After 1980, however, women were more likely to be elected if they were in the top 10% of publications. Extending these analyses to mathematics and psychology, Card et al. (2023) show that women with similar publications were less likely to be elected to the National Academy of Science and the American Academy of Arts and Science in the 1960, 70, and 80s but are currently more likely to be elected. While these results indicate that conditions for women have improved, survey data suggest that children of college-educated parents (as a proxy for parental SES) are overrepresented among PhDs, especially at top economics programs (Stansbury and Schultz 2023) and among professors (Morgan et al. 2022). Examining changes in the distribution of award-winning researchers across universities since the 1950s, Freeman et al. (2024) document declining concentration in all fields except economics. We complement these findings by investigating the early influence of SES on peer recognition across 12 disciplines, and by showing that these results are robust to differences in the number and the quality of publications, elite degrees, and other traits of scientists.

Our results also contribute to analyses of intergenerational mobility (e.g., Black and Devereaux 2011; Chetty et al. 2014) and, more specifically, to the growing evidence on the role of universities in diversifying the professional elite. Analyzing intergenerational mobility among students at U.S. colleges, Chetty et al. (2017) show that elite college degrees offer an effective path to upward mobility for children from low-income families. Children from the bottom 20% of the income distribution are, however, 77 times less likely to attend elite colleges than children from the top 1%. Exploiting idiosyncratic variation in admissions decisions for waitlisted applicants, Chetty, Deming, and Friedman (2023) show that attending an Ivy-Plus university (Ivy League, plus Stanford, MIT, Duke, and Chicago) instead of a public flagship increases students' chances of reaching the top 1% of the earnings distribution by 60%, nearly doubles their chances of attending an elite graduate school, and triples their chances of working at a prestigious firm. Our analysis complements these findings by investigating the influence of elite education one century earlier, at a critical junction in U.S. history. Analyzing old boys' clubs at Harvard in the 1920s, Michelman, Price and Zimmerman (2022) show that students from prestigious private schools were overrepresented among club members while academic high achievers were largely absent; upon graduation, members earned 32% more. We add to this literature by investigating how elite colleges influence peer recognition and stardom, a key marker of success that is critical to hiring and promotions, yet poorly understood.

I. DATA

"There were two awards whose value was apparent to all in the 1920s and 1930s: a 'star' in the *American Men of Science*, and the supreme accolade, the Nobel Prize." (Rossiter 1982, p. 289)

Data cover all 9,554 entries in the *American Men of Science* (MoS 1921), including stars and other scientists. First published in 1906 by James McKeen Cattell, the MoS is an exceptionally comprehensive source of biographical information for male and female scientists in the United States and Canada. Cattell collected these data originally for his own research on the psychology of intelligence. Born into a wealthy Pennsylvania family, Cattell earned his PhD in Leipzig, Germany, and became the first American to publish a dissertation in psychology. Between 1894 and his death in 1944, Cattell served as the editor of *Science*.

Cattell used this expertise to establish a comprehensive compendium of scientists whose work "contributed to the advancement of pure science." To collect his data, Cattell started with the membership of scientific societies:

"The National Academy of Sciences, Fellows of the American Association for the Advancement of Science, the American Society of Naturalists, the Association of American Anatomists, the Association of American Geographers, the Association of American Physicians, the American Association of Pathologists and Bacteriologists, the Astronomical and Astrophysical Society of America, the Geological Society of America, the American Mathematical Society, fellows of the American Ornithologists' Union, the American Philosophical Association, the American Physical Society, the American Psychological Association, the American Society of Bacteriologists, the Society for the Promotion of Agricultural Science, the Society for Experimental Biology and Medicine, the Society for Horticultural Science, the Society for Plant Morphology and Physiology, and the American Society of Zoologists." (Preface to the first edition, 1906, p. v.)

To these membership lists, he added names from "catalogues of institutions of learning," along with the authors of articles in scientific journals. In addition, Cattell printed requests for names omitted in *Science*, *The Popular Science Monthly,* and in *The Nation* (Cattell 1906, p. vi).

As a result of this thorough process, the MoS is "tolerably complete for those in North America who have carried on research work in the natural and exact sciences" (Cattell 1921, p.v), including a total of 9,554 scientists—a response rate of 95% relative to 10,000 requests. With the publication of his data, Cattell hoped to provide the "chief service" to the profession to "make men of science acquainted with one another and with one another's work" (Cattell 1921).

Entries in the MoS include the scientist's full name, date and place of birth, discipline, university education, and employment history. Simon Flexner, for example, was born in "Louisville, Ky, March 25, 1863" (Figure A1). Information on the exact birth date and place is available for 97% of scientists. This information is invaluable for matching scientists with the census and allows us to separate U.S.-born from foreign-born scientists. 8,146 of 9,554 scientists in the MoS (1921, 85%) are U.S.-born. The oldest scientist in the MoS (1921) is Francis H. Smith, a physicist born in 1829; the youngest is the chemist William A. Noyes Jr, born in 1898.

We focus on male scientists because women are difficult to match with their childhood home. Using historical gender frequencies in the US Social Security Administration Records between 1880 and 2011, Python's *gender-detector* package identifies 7,791 U.S.-born scientists (96% of all U.S.-born scientists) as male.

Stars

To identify star scientists, Cattell asked their peers to choose "leading scientific men arranged in the order of merit" (Cattell 1906, p. 699). For instance, Simon Flexner's entry shows a * next to "Pathology" (Figure A1). Among a total of 9,554 scientists in the MoS (1921), 1,322 (13.8%) were stars. Among 7,791 U.S.-born males in the MoS, 1,112 (14.3%) were stars.

Criteria of merit emphasize "contributions to the advancement of science, primarily by research." In addition, contributions to "teaching, administration, editing, the compilation of textbooks, etc., should be considered" (*Science* 1910, p. 635). For the MoS (1906 and 1910), Cattell asked 10 leading scientists in each of the 12 disciplines to rank their peers and attached a star to the top 1,000 entries.

To identify stars for the MoS (1921), Cattell asked stars in the 1906 and 1910 editions to nominate up to ten people whom they considered leaders in their discipline. Next, he asked scientists who were nominated at least twice to create a shortlist of ten names. From this list, Cattell extracted the top quartile for the 12 disciplines. This created a list of electors, ranging from just 20 scientists in anthropology to 100 in psychology and 175 in chemistry. Electors then marked scientists in the top 50% with one checkmark and those in the top 5% with a double checkmark. Cattell added these votes to determine stars.

While the MoS (1921) does not report the year when a scientist first became a star, we collect this information from *Scientists starred, 1903-1943, in "American Men of Science* (Visher 1947). In addition to the 1,322 scientists marked by *, another 443 scientists in the MoS (1921) became stars after 1921; we treat them as non-stars in the main specifications.[3]

[3] In the late 1930s, this system of adding stars began to break down, partly because Cattell found it difficult to take away stars from older scientists whose work no longer put them in the top 1,000 (Rossiter 1982, p. 289). When his son Jacques took over the MoS in 1943, he abolished the system of stars.

Using Text Analysis to Assign Scientists to 12 Disciplines

Cattell asked scientists within 12 disciplines to elect stars; these disciplines, however, are only observable for stars but not for other scientists. Visher (1947) reports the names (and disciplines) of 2,605 stars elected between 1906 and 1939.[4] We match stars in Visher (1947) to the MoS (1921) to create training data for a nearest centroid matching algorithm. We then apply the trained algorithm to the research topics of the remaining scientists in the MoS to match each non-star with a unique discipline.

First, using the full name and date of birth, we match 1,782 stars scientists in the MoS (1921) with Visher and assign the discipline reported in Visher to these stars.[5] Next, we assign 4,342 scientists who list one of the 12 disciplines as their subject in the MoS to the respective discipline. Flexner, for example, is one of 176 scientists in "pathology." We assign all 176 scientists to the discipline of "pathology."[6] Finally, we exploit unique links between 89 subjects and the 12 disciplines to assign 1,487 non-star scientists uniquely to one of the 12 disciplines. Charles Dana, for example, is a star in "pathology" and lists "nervous and mental diseases" as his subjects in the MoS. Since Dana is the only star in "nervous and mental disease," this subject is uniquely linked with pathology. Using this one-to-one link, we assign all scientists who list their subject as "nervous and mental disease" to the discipline of "pathology."

This three-step process creates a training data set of 7,611 scientists matched to 12 disciplines. We use the text that describes the research of these 7,611 scientists to train a nearest centroid classification algorithm that allows us to assign the remaining 1,943 scientists to disciplines. Flexner, for example, describes his research as

"bacteriology; pathology of toxalbumin intoxication; terminal infection; snake venom; histological alterations of eytoxinic intoxication; etiology of dysentery; serum therapy of epidemic cerebrospinal meningitis; etiology and pathology of infantile paralysis; lethargic encephalitis."

[4] 116 scientists are stars in more than one discipline, we assign them to the discipline in which they ranked highest.
[5] 1,322 stars in Visher are stars in the MoS (1921); 443 other scientists became stars after 1921.
[6] Scientists in the MoS (1921) report 385 subjects. All scientists in the MoS report a primary subject; 988 scientists report one or more additional, secondary subject(s). Disciplines range from extremely broad (such as chemistry, with 1,504 scientists, physics 679 scientists, and mathematics 604 scientists), to extremely narrow (such as Water Supply and Child Hygiene, with 1 scientist each). 155 of 324 primary subjects in the MoS have just one scientist, making it impractical to use disciplines alone to classify scientists.

Research topics are available for 93.7% (8,951 of 9,554 scientists in the MoS (1921) and 90.8% (1,764) of the remaining 1,943 scientists.

Methodologically, we use a GloVe model (pre-trained on 300-dimensional word vectors from Wikipedia 2014 and Gigaword 5, Pennington et al. 2014) to create vector representations of scientists' research subjects and topics. With this vectorization, a scientist is represented by the average of the GloVe vectors of the words that describe their research. We use the subject and research topics of the 7,611 scientists whom we can assign to a unique discipline to train a nearest centroid classifier model (using Euclidian distance as the distance metric) and apply the trained algorithm to assign the 1,943 remaining scientists uniquely to the 12 disciplines.

A key advantage of GloVe embeddings and the nearest centroid metric is that they can recognize similar words that lie outside the average human's vocabulary but may be used by scientists. For example, the target words for "frog" include "toad," along with scientific terms such as eleutherodactulus and leptodactylidae, which would be difficult to recognize in manual assessments. Word clouds represent the most frequent words that describe the research topics in each discipline (Figure B2). As a robustness test of the algorithm, we compare the algorithmic assignment with manual assignments for a random sample of 200 scientists (Appendix B).

With 1,984 and 1,480 scientists, chemistry and physics are the largest disciplines (Figure 1, Panel A), followed by botany (1,242), zoology (987), pathology (855), geology (841), mathematics (622), and psychology (520). The four smallest disciplines are physiology (370), anatomy (286), astronomy (235), and anthropology (132). Astronomy has the largest share of stars: 28.1% of all scientists who are astronomers are stars (Figure 1, Panel A). Astronomers are also among the oldest when they first become stars, at a median of 45 years and a range of 31 to 66 years. The oldest scientist to become a star is a biologist, John Gulick, who studied "*Achatinella"* (a tree snail) and became a star in 1910, at the age of 78.

Physics, mathematics, and physiology have the lowest median age of becoming a star, with 40, 39, and 39 years, respectively. The youngest person to become a star is Irwin Priest, a physicist who studied *"Interferometry; wave lengths; flexure; elasticity, fatigue, and effects of straining iron and steel"* and became a star in 1910 at the age of 24. The second youngest star is the mathematician Edwin Wilson, who studied *"multiple algebra; mechanic; advanced calculus; relativity; aeronautics"* and became a star in 1906 at the age of 25.

Linking Scientists with Census Records

We use machine-learning algorithms to match all 7,791 U.S.-born male scientists in the MoS (1921), with 16.5 million U.S.-born entries in the census of 1870, 21.8 million in 1880, and 33 million in 1900. Methodologically, we adopt machine learning methods that Feigenbaum (2016) developed to link individuals across census waves to match scientists with the census (see Appendix C for details).[7]

First, we create a training data set by hand-matching 2,000 scientists with the census of 1880 and hand-matching 1,000 scientists, each with the census of 1870 and 1900. Next, we create a script that defines potential matches as census records whose first and last names are within a 0.2 Jaro-Winkler string distance of the scientist's name and born in the same state within three years of the scientist's year of birth. We manually review all potential matches, select the best match for scientists with multiple matches, and classify other MoS-census pairs as false matches.

Using 1,000 scientist-census matches for 1880, half of the training sample for 1880, we fine-tune the matching parameters from Feigenbaum (2016).[8] Specifically, we estimate a probit model for a matching score between 0 and 1 for each scientist-census pair, where values closer to 1 indicate greater similarity. Matching variables reflect similarity in names, birth years, and characteristics of other potential matches.[9] For each scientist, we choose census records whose match probability is above a minimum threshold in absolute terms and relative to second-best candidates. We fine-tune this step to maximize a combination of recall and precision. Recall is the True Positive Rate (TPR), the share of correctly identified matches over the set of possible matches: TPR = True Positives/ (True Positives + False Negatives). Precision, or the Positive Predicted Value (PPV), is the share of correctly identified matches over the number of scientist-census pairs classified as matches, PPV= True Positives/ (True Positives + False Positives). We set a minimum of 90% precision to train and validate the algorithm.

[7] Linking birth certificates with schooling records for children born in Florida 1992-2002, Autor et al. (2019) find that relative to their sisters, boys from disadvantaged families have higher rates of disciplinary problems, lower achievement scores, and fewer high-school completions, which suggests that childhood SES matters more for men.

[8] Feigenbaum (2016) uses the matching procedure to match children in the 1915 Iowa State Census to their adult-selves in the 1940 Federal Census; he matches 57.4% of children with their adult self.

[9] To measure distance in names, we use Jaro-Winkler and Soundex distances and indicators for a match between the first or last letter of the first, middle, and last name. To measure distances in birth years, we use dummies for births that are one, two or three years apart. Additional matching variable include an indicator for an exact name match, the total number of exact name matches in the sample, and the number (in levels and squared) of census candidates.

The remainder of the hand-matched training sample allows us to evaluate the out-of-sample performance of our matching algorithm. 1,000 matches between the MoS (1921) and the census of 1870, 1880 and 1900 respectively. Out-of-sample performance is high, with 81-87% for recall and 83-88% for precision (Table C3).

Applying this machine-learning algorithm to the MoS (1921), we match 6,104 U.S.-born male scientists (78.4%) with at least one census wave and observe 4,866 (67.4%) of scientists in their childhood home in at least one census year.[10] 945 stars (71.5%) matched scientists and 699 (52.9%) of scientists whom we observe as children are stars. Matched scientists are distributed similarly across fields (Figure 1, Panel B) to the full sample (Figure 1, Panel A), and matching rates are consistent across fields, ranging from 57.3% of U.S.-born anthropologists to 62.4% of chemists (Appendix Figure C1). Matched and unmatched star and other scientists are also similar in terms of age and education (Table A1, Panel B, Figures A3-A5).

Measuring Childhood SES by Matching Scientists with their Census Records

We use the occupation of the scientist's father to measure the scientist's childhood SES. Father's occupation is available for 4,067 of 4,866 (84%) census-matched male U.S.-born scientists, including 1,142 of 1,334 (85.6%) scientists who were minors in the census of 1870, 2,274 of 2,566 (88.6%) in 1880, and 1,182 of 1,644 (71.9%) in 1900.[11] For scientists whom we observe as minors in more than one census year (678 scientists each in 1870 and 1880), we use the father's occupation in the 1870 census.[12] By this process, we observe the childhood SES of 4,067 scientists, 67% of all U.S.-born males who were minors in a census year, and 14.6% of stars.

OCCSCORE is an income score measuring the relative economic standing of occupations using the median total income—in hundreds of dollars—for persons with positive income in 1950. As an alternative measure, the ERSCORE counts the percentage of persons in occupations having lower standardized median earnings than the respondent's occupation. The EDSCORE

[10] We match 1,334 scientists (57.5%) who were minors in 1870 with the census of 1870, 2,566 (65.6%) with 1880, and 1,644 (64.4%) with 1900. 678 scientists are observed in both the 1870 and 1880 census (Table 1, Panel B).

[11] 50 scientists in 1870, 69 in 1880 and 291 in 1900 live with a father who is retired, unemployed, or has an unclassified occupation. Nine fathers pursue an occupation in the armed forces, so that ERSCORE, EDSCORE and Siegel are not available for them. Another 372 (7.5%) of male U.S.-born scientists live with a single mother, including 98 scientists in 1870 (7.3%) ,153 (6%) in 1880, and 131 (7.6%) in 1900. Since few mothers work (1% in 1870, 14% in 1880 and 4% in 1900), we cannot observe parental SES consistently for these scientists.

[12] We match 678 scientists to both the 1870 and 1880 census and observe the father's OCCSCORE in both years for 531 scientists; for these scientists we define high-SES using the earlier wave in 1870. Since we focus on scientists below 18 when we observe them in the census, there is no overlap between children matched in 1880 and 1900.

captures the share of individuals in each occupation with one or more years of college education in 1950. The Duncan Socioeconomic Index, SEI, is a weighted sum of occupational income and education based on a 1947 survey by the National Opinion Research Center. The Siegel (1971) prestige score, PRESGL, uses survey responses on the "general standing" and "social standing" of occupations from the National Opinion Research Center in the 1960s.

As an alternative to these composite measures of SES, we create a simple, directly observable measure of SES using the presence of live-in servants. Servants are observable in the census of 1880 and 1900. We construct this variable for 2,516 scientists in 1880 and 1,671 in 1900. Encouragingly, the presence of servants is positively correlated with the OCCSCORE, even though the two measures of SES are methodologically distinct.

In addition, we create a wealth-based measure that is available for the 1870 census measured as personal property (*persprop* in IPUMS): "the contemporary dollar value of all stocks, bonds, mortgages, notes, livestock, plate, jewels, and furniture owned by the respondent."

Matching Scientists with Publications and Citations

To measure scientific productivity, we match scientists with their publications and citations from Microsoft Academic Graph (MAG, Sinha et al. 2015). MAG was updated weekly until December 2021; we use the version from August 20, 2020. To perform the matching, we restrict the data to authors with at least one English-language publication between 1900 and 1970. We match scientists in the MoS (1921) with *authorids* in the MAG, using first and last names, as well as middle initials. Using information on birth years from the MoS, we further restrict matched publications to those published when the scientist was between 18 and 80 years old. To measure achievements *before* becoming a star, we restrict the analyses to articles and books published before 1921. Our final data consists of 98,076 publications matched to 6,592 scientists. 5,485 U.S.-born scientists (70% of all U.S.-born scientists) and 3,423 (71.3%) U.S.-born scientists matched with their childhood home have at least one publication before 1921. Conditional on having at least one publication, the median U.S.-born scientist has 8 publications by 1921, and the median scientists matched with his childhood home has 4 publications by 1921.

We construct four complementary measures for the quality of publications. Three citations-based measures distinguish highly cited authors in the top 10, 5, and 1% of their disciplines. Conditional on having at least 1 publication, the median matched scientist has 1.5 citations per paper, with an average of 9.4. To complement citations-based measures, we create an indicator for papers in *Science*, which has consistently been a top journal across disciplines. 763 (15.7%) of the matched scientists have at least one publication in *Science.*

Matching Scientists with Patents

To measure scientific productivity, we match scientists with their patents using Google Patents. We construct a matching algorithm that matches the full name of scientists with the names of inventors on 2,748,078 successful patent applications filed between 1880 and 1970. Requiring precise matching of names for computing efficiency, we match 6,277 scientists to 48,345 successful patent applications filed between 1880 and 1970.

We use co-inventor connections until 1921 to investigate whether scientists who became stars in 1921 had collaborated with an existing star. A total of 22,635 patents by 2,170 scientists in our matched data were filed by 1921. Among 4,067 scientists who we observe as minors in their childhood home, 915 scientists (22.5%) have at least one patent by 1921. 408 scientists (10.0%) have patents with co-inventors, and only 34 (0.8%) have at least one patent with a star.

Academic versus Industry Scientists

To separate academic from industry scientists, we exploit detailed information on scientists' career histories in the MoS. We define academic scientists as the 7,622 scientists in the MoS (1921) who held an academic position, such as a lecturer, assistant professor, or associate professor, at least once in their career. The remaining 1,932 scientists are industry scientists. Flexner, for example, is an academic scientist because he was an associate professor at Johns Hopkins from 1896 to 1899 (Figure A1), while William de Chastignier Ravenel, a scientist at the U.S. Fish Commission who never worked in academia, is an industry scientist. 81.1% of the 4,081 scientists whom we observe as children are academics, while just 18.9% work in industry.

Elite University Degrees

Universities have been shown to help shape generational persistence (e.g., Chetty et al. 2020, Michelmann et al. 2022) and may have an even greater impact on science. We explore this channel by collecting information on college attendance and degrees from the MoS.[13] 7,964 scientists (83.4% of all scientists) have graduate degrees, including M.A., M.D.s or PhDs, and 6,831 U.S.-born scientists (83.9%) have graduate degrees. Data on undergraduate degrees are available for 7,844 scientists (82.1% of all scientists) and 6,861 U.S.-born males (84.2%). We create unique identifiers for universities by first cleaning acronyms and creating a crosswalk between historical institutions in the MoS (1921) and 1,060 unique contemporary institutions.

To define "elite" education, we implement Chetty et al.'s (2017) definition of Ivy-Plus universities, which includes eight Ivies (Brown, Columbia, Cornell, Dartmouth, Harvard, Penn, and Yale), along with Chicago, Duke, MIT, and Stanford. 23.5% of scientists (and 31.7% of stars) in the MoS 1921 earn at least one undergraduate degree from an Ivy-Plus institution, and 40.1% of scientists (and 46.4% of stars) earn at least one graduate degree from an Ivy-Plus institution.

II. Who Becomes a Scientist?

First, we examine whether scientists were disproportionately drawn from high-SES families. Specifically, we compare the occupational status of scientists' fathers for 4,866 U.S.-born male scientists in the MoS (1921) with a stratified 5% sample of U.S.-born males of the same age in the same census waves. Existing research has shown that children from high-SES families are overrepresented among inventors today (e.g., Bell et al. 2019; Aghion et al. 2018).

Children from High-SES Families are Overrepresented in Science

Linked scientist-census data show that patterns of underrepresentation today were already present in 1921. Children of attorneys, physicians, and clergymen are overrepresented in science, while children of farmers and other low-income occupations are underrepresented. 6.9% of scientists whom we observe as children in the census of 1880 are the sons of clergy, compared with 0.6% of boys of the same age in the population (Figure 2). Similarly, 4.9% of scientists are the sons of physicians (0.7% of boys in the population), and 3.3% are the sons of attorneys (compared with 0.5% of boys in the population).

[13] We separate summer programs, exchange programs, fellowships, and other non-degree programs from degrees. 8.2% of scientists (and 7.5% of stars), earn two or more undergraduate degrees and 38.6% of scientists (and 49.2% of stars) earn two or more graduate degrees (excluding honorary degrees).

Notably, just 34.4% of scientists whom we observe as children in the 1880 census are the children of farmers, compared with 55.6% of boys in 1880. Even more strikingly, 0.8% of scientists are the sons of farm laborers, compared with 6.1% in the population.

Patterns of underrepresentation are remarkably stable over time (Table 2, Panel A). For instance, 45.6%, 55.6% and 56.2% of boys in the census of 1870, 1880 and 1900 are the sons of farmers, compared with just 31.1, 34.4%, and 30.3% of future scientists (see Figure 3 for the census of 1870 and Figure 4 for the census of 1900).

Using the occupational income rank (OCCSCORE) of a person's father as a measure of childhood SES, we find that the median scientist is drawn from the 78th percentile of income in 1880 (with an OCCSCORE=24, equivalent to the rank of a salesclerk or carpenter), while the median boy of the same age in the population is the son of a farmer (OCCSCORE = 14). For scientists in the census of 1870 and 1900, the median OCCSCORE is 24, compared with 14 in the population (Figures 3 and 4).

Scientists are also more likely to have grown up in households with servants: 8.6% of scientists who were children in the census of 1880 lived in households with servants, compared with just 0.5% of the population (Table 2, Panel B). 24.1% of scientists who were children in 1900 lived in households with servants, compared with just 1.4% of the population. Capturing differences in parental education and occupational prestige, alternative measures of SES imply even larger differences in participation (Table 2, Panel B).

Taken together, these results suggest that patterns of inequality that impact innovation and science today (e.g., Bell et al. 2019; Aghion et al. 2018) already existed in the early 20th century, affecting children born as early as the 1860s.

III. WHO BECOMES A STAR?

Beyond participation, SES may impact a person's professional success through selection, inequities in educational opportunities, peer recognition, and many other forces. If barriers to entry exist, scientists from low-SES backgrounds may be positively selected, increasing their likelihood of becoming stars. If educational opportunities are unequal, however, children from low-SES families may be less productive, reducing a person's chances of becoming a star.

Moreover, if recognition for achievements is conferred primarily through social networks, individuals from low-SES families may receive less recognition for the same work.

Children from High-SES Families are More Likely to Become Stars

Comparing how stars and other scientists are distributed across the spectrum of childhood SES, we find that the children of managers, clergy, and attorneys are overrepresented among stars relative to other scientists, as well as relative to the population (Figure 2): 19.9% of stars whom we observe as children in 1880 are the children of managers, compared with just 16.3% of scientists and 5.6% of boys of the same age in the population; 8.7% of stars are children of clergy (compared with 6.9% and 0.6%) and 4.7% are children of attorneys (compared with 3.3% and 0.5%).

At the same time, the children of farmers and farm laborers are underrepresented among stars (Figures 2-4, Panel C). Just 29.0% of stars we observe as children in 1880 are the children of farmers, compared with 34.4% of scientists and 55.6% of the population. 0.0% of stars are the children of farm laborers, compared with 0.8% of scientists and 6.1% of boys in the population.

Comparing the fathers of stars, scientists, and the population, we find that stars had fathers with the highest OCCSCORE, with a median of 25 in 1880 (equivalent to a kindred worker), compared with 24 (carpenter) across all scientists and 14 (farmer) for the population (Figure 2).

Notably, SES seems to matter much more for scientists whom we observe when they are young, below the age of 40 (Figure 4, for scientists whom we observe as children in the census of 1900). At that age, observers have less information about the "quality" of the scientist, and they may put greater weight on traits that are associated with a scientist's class, such as their social assuredness. By comparison, parental SES appears to matter much less for becoming a star when scientists are old, above the age of 50, when observers have more information on quality (Figure 3, for scientists whom we observe as children in the census of 1870).

Logit Estimates of a Scientist's Odds of Becoming a Star

To evaluate differences in parental SES more systematically, we estimate maximum likelihood logit models of the probability of becoming a star, controlling for the scientist's age, discipline, and the year when we observe him in the U.S. census:

$$P(star) = \left(1 + e^{-(\alpha + \beta SES_{high\,SES} + \delta_c + \delta_d + \gamma age)}\right)^{-1} \qquad (1)$$

where $SES_{high\,SES}$ is an indicator for scientists whose father was in the upper half of the distribution of SES in a given census year, δ_c is a vector of census year fixed effects for 1870 and 1900, with 1880 as the excluded category, δ_d is a vector of fixed effects for 11 of the 12 scientific disciplines (using chemistry, the largest discipline, as the excluded category), *age* is the scientist's age in 1921, and α is an intercept. The coefficient $\hat{\beta}$ measures log odds. Converting it using the exponential function $e^{\hat{\beta}}$ captures the odds ratio of being a star for scientists from high-SES families compared with other, low-SES scientists. We calculate the additional odds for a high-SES scientist of becoming a star as

$$\%\,\Delta \text{ in odds} = e^{\hat{\beta}} - 1 \qquad (2)$$

Estimates for SES are consistently positive and significant (Table 3). Scientists who grew up above the median OCCSCORE of 24 (ranked above a clergy, carpenter, or salesclerks) have 38.7% (= $e^{0.327} - 1$) greater odds of being stars compared with other scientists (Table 3, Column 1). Scientists who grew up in a household with live-in servants (14.6% of all scientists) had 75.7% higher odds of being stars (Table 3, Column 6).

Converting estimates to probabilities confirms the influence of SES. Table 3 reports probabilities evaluated at the baseline averages (specified in Table D1): A 46-year-old chemist scientist who lived in a high-OCCSCORE home in the 1880 census has a 33.5% greater probability of being a star compared with a scientist of the same age from a low-OCCSCORE family. A 46-year-old scientist from a household with live-in servants in 1880 has a 63.9% higher probability of being a star. Since probabilities must be evaluated at specific values of the explanatory variables, we report coefficients in odds and convert only key values into probabilities. Appendix D provides details on all conversions.

Alternative measures of SES suggest an even stronger link between childhood SES and peer recognition (Table 3, columns 3 to 5). Scientists from the top of the ERSCORE (an alternative measure for the occupational income rank) have 35.0% higher odds of becoming stars. Estimates for measures of occupational prestige yield similar results, with an estimate of 44.4% for the Duncan SEI and 53.0% for the SIEGEL.

Scientists from families in the top half of the EDSCORE have 50.3% higher odds of becoming stars (Table 3, Column 2). Recent surveys of PhD recipients show that children of college-educated parents are overrepresented among U.S.-born PhDs: 50% of U.S.-born PhD students had a parent with a graduate degree, and 12% had a parent with a PhD (Stansbury and Schultz 2023). Moreover, a 2017-20 survey of 7.204 U.S.-based tenure track faculty across eight disciplines in STEM, social science, and the humanities shows that faculty are up to 25 times more likely to have a parent with a PhD (Morgan et al. 2022). Our results indicate that parents' education already mattered in the early 20th century and that it mattered for peer recognition, conditional on participation.

IV. Why are Children from High-SES Families More Likely to Be Stars?

In this section, we investigate mechanisms by which a scientist's childhood SES may influence their odds of becoming a star, starting with publications. We also examine differences across disciplines and between academia vs. industry. We explore the influence of connections to existing stars, degrees at elite universities, family size, birth order, foreign-born parents, illiterate mothers, links with the father's occupation, and personal wealth.

Are High-SES Scientists Stars because they Publish More and Better Papers?

People from high-SES families may become stars because SES is associated with better health, cognitive, and socioemotional outcomes (e.g., see Bradley and Corwyn 2002), enabling them to produce more and better papers.[14] To investigate this mechanism, we re-estimate equation (1) controlling for the number and the quality of publications:

$$P(star) = \left(1 + e^{-(\alpha + \beta SES_{high\,SES} + \theta pubs_d + \delta_c + \delta_d + \gamma_{age})}\right)^{-1} \qquad (3)$$

where $pubs_d$ counts a scientist's publications before 1921 and citations to these publications. In the baseline specification, publications and citations are transformed by inverse hyperbolic sine (*asinh*) to account for the large share of scientists (29%) without publications and the large share

[14] In politics, this mechanism is what Dal Bó, et al. (2017) call "exclusive meritocracy:" "politics select the competent, which makes the political class (accidentally) elitist."

of publications (56%) without citations.[15] Alternative specifications include a logarithmic specification (adding 0.01 to the count of publications and citations), indicators for scientists in the top 1, 5, or 10% in the distribution of publications and citations within their discipline *d*, as well as publications in the prestigious journal *Science*.

Logit estimates of equation (3) show that publications and citations increase the odds of becoming a star, but they cannot explain the influence of SES. Controlling for publications and citations, scientists from the top half of the distribution of OCCSCORE have 38.3% higher odds of becoming a star (Table 4, Panel A, column 2), just 0.4% less compared with 38.7% without controls for publications (column 1). Estimates with alternative controls for scientists in the top 1, 5, and 10% of publications and citations imply 37.7-40.6% higher odds of stardom for scientists from high-SES families (columns 3 to 5); estimates with controls for papers in *Science* imply 35.4% higher odds (column 7).

Regressions with servants as an alternative measure for SES confirm these results (Table 4, Panel B). Scientists who grew up in households with servants have 55.8% higher odds of becoming a star than scientists without servants (Table 4, Panel B, column 2).

Does the Influence of SES Vary across Disciplines?

SES may matter more in some disciplines, depending, for example, on differences in status bias or access to education.[16] A pathologist from a low-income family, for example, may have found it more difficult to get top of the line training in the early 20th century, when most medical education was for-profit (Duffy 2011, p. 273), and only a small number of elite universities in the Northeast trained students in the cutting-edge European model, teaching two years of laboratory sciences before progressing to clinical training in hospital wards (Duffy 2011, p. 271). Under these conditions, children from high-SES families may have become stars because they were better educated, enabling them to publish more and better papers.

[15] $\text{asinh}(x) = \ln(x + \sqrt{1+x^2})$. The transformation *asinh* is well-defined at zero as $\text{asinh}(0) = 0$. For $x > 2$, $\text{asinh}(x) \approx \ln(x) + \ln(2)$. Investigating log transformations of the outcome variable, Chen and Roth (2024) argue that ATEs for the asinh (and other log-like transformations) should not be interpreted as approximating percentage effects, since log-like transformations depend on the units of the outcome variable, unlike a percentage. While we use the asinh to transform the explanatory variable, we also present alternative measures without the asinh.

[16] Surveys of PhD recipients today show that just 13% of U.S.-born PhD students in economics were first-generation college students (defined as having no parent with a college degree), compared with 49% in education, 19% in math, 22% in the physical sciences, and 26% across all disciplines (Stansbury and Schultz 2023, p. 2010).

Class differences in access to participation may be particularly salient in disciplines in which access to early childhood education is critical for success. For instance, the National Research Council's (2009, pp.1-2) report on *Early Childhood Education in Mathematics* emphasizes the importance of early childhood education and concludes that

"although virtually all young children have the ability to learn and become competent in mathematics, for most the potential to learn mathematics in the early years of school is not currently realized. This stems from a lack of opportunities to learn mathematics either in early childhood settings or through everyday experiences in homes and in communities. This is particularly the case for economically disadvantaged children, who start out behind in mathematics and will remain so without extensive, high-quality mathematics instruction."

To investigate variation across disciplines, we estimate equation (1) within the 12 disciplines:

$$P_d(star) = \left(1 + e^{-(\alpha + \beta SES_{high\,SES} + \delta_c + \gamma_{age})}\right)^{-1} \qquad (4)$$

Estimates reveal much variation across disciplines. SES matters most in geology, a field in which high-SES scientists have 117.1% higher odds of becoming stars (Table 5), followed by psychology (109.0%), pathology (103.6%), botany (89.7%), and math (73.5%).

In nearly all fields, controlling for publications intensifies the link between high-SES and becoming a star: Pathologists from high-SES families have nearly three times higher odds of becoming a star (with an estimate of 1.379, Table 5, significant at 1%), implying a 297.1% increase in odds. In geology, high-SES scientists have 207.7% higher odds of stardom (with an estimate of 1.124 significant at 1%). In psychology, they have 135.4% higher odds (0.856, significant at 10%), and in botany, they have 104.0% higher odds (0.713, significant at 10%). Mathematics is the only field in which the link between SES and stardom operates through publications. Without controls for publications, high-SES mathematician scientists have 73.5% higher odds of becoming a star. With controls for publications, the coefficient on SES becomes small and is no longer statistically significant. This result is consistent with inequities in access to early childhood education, which puts children from lower SES families at a disadvantage (National Research Council 2009).

Does SES Matter More in Academia or Industry?

In addition to disciplines, the types of jobs that scientists hold may influence whether they are perceived as stars, irrespective of publications. Specifically, scientists may attach more prestige to university employment if only because they were trained at universities and advisors push PhD students to strive for university appointments. Moreover, if academia is more elitist, childhood SES may matter more in academia relative to industry.

Notably, industry scientists are more likely to come from high-SES families than academic scientists, suggesting that academia may, in fact, be more egalitarian than industry. The father of the median industry scientists has an OCCSCORE of 24, compared with 26.5 for academic scientists (Figure 6, Panels A and B). At a more granular level, 5.9% of industry scientists are the children of lawyers and judges, compared with 5.0% of academic scientists. Most strikingly, the children of farmers, who make up 34.2% of academic scientists, are underrepresented among industry scientists, with a share of 25.1%.

If academia is more egalitarian than industry, childhood SES may matter more in industry. For instance, it may be harder for a child from a low-SES family to become an industry scientist because lucrative jobs may be transferred through connections (e.g., San 2023). To control for such differences, we first add an indicator for *industry* scientists to equation (3):

$$P(star) = (1 + e^{-(\alpha + \beta SES_{high\,SES} + \lambda industry + \theta pubs + \delta_c + \delta_d + \gamma_{age})})^{-1} \qquad (5)$$

Logit estimates imply that industry scientists are 66.6% less likely to be stars than academic scientists (Table 6, column 2, significant at 1%, without controls for publications). Controlling for publications increases this industry penalty to 68.8% (column 3, significant at 1%). Controlling for industry employment and publications, scientists from high-SES families have 46.9% higher odds of being stars (column 3, significant at 1%), and scientists from households with servants have 66.7% higher odds of being stars (Table A2, column 3, significant at 1%).

Estimating the baseline separately for industry and academic scientists confirms that the link between childhood SES and stardom is stronger in industry. Controlling for publications and citations, an industry scientist from a high-SES family has 72.7% higher odds of becoming a star (Table A3, column 4) compared with 45.9% in academia (Table A3, column 2).

Do Elite Universities Mitigate or Exacerbate Inequality in Sciences?

Next, we investigate whether elite universities mitigate inequality or whether they make it worse. Chetty et al. (2017) match students at U.S. colleges between 1999 and 2013 with their parents' tax records to evaluate colleges' effectiveness in encouraging upward mobility. They find that elite Ivy-Plus universities—Brown, Columbia, Cornell, Chicago, Dartmouth, Duke, Harvard, MIT, Pennsylvania, Princeton, Stanford, and Yale—succeed in moving admitted low-income children up in the income distribution but fail at training enough of these students. Elite colleges may, however, amplify the importance of socioeconomic inequality in science if elite colleges are more likely to teach cutting-edge materials (Biasi and Ma 2023), granting their graduates a head start. Moreover, elite colleges may amplify inequality if they create social networks that disproportionately benefit people of higher SES (Michelman et al. 2022).

To investigate the role of elite colleges in peer recognition, we add controls for Ivy-Plus degrees to equation (5). These estimates confirm that graduates from elite colleges have significantly higher odds of becoming stars, even controlling for publications and academic jobs. Scientists with elite undergraduate degrees have 41.5% higher odds of becoming stars (with a coefficient of 0.347, Table 6, column 4 significant at 1%), and scientists with elite graduate degrees have 48.9% higher odds (with an estimate of 0.398, column 5, significant at 1%).

Controlling for elite undergraduate degrees reduces the estimate for high-SES more than any other control: Controlling for elite undergraduate degrees (in addition to publications and academic jobs), high-SES scientists have 43.3% higher odds of becoming stars (Table 6, column 4, and Figure 5, Panel A), the largest reduction relative to the baseline of 46.9% (column 3). Similarly, scientists who grew up in households with servants have 58.1% higher odds (Table A2, column 4 and Figure 5, Panel B), a large reduction relative to the baseline of 66.7% (Table A2, column 3). Controlling for elite graduate degrees leaves the estimates at 44.9% for high-SES (Table 6, column 5) and at 66.6% for scientists with servants (Table A2, column 5).

Are First-Born Children More Likely to be Stars?

To further investigate the channels by which SES influences stardom, we extend the analyses to control for family size (measured by the number of siblings) and birth order (measured by being the oldest living child). First-born children may be more likely to become scientists (and stars) because they benefit from higher level of prenatal investment (Buckles and Kolka 2014) or more quality time with parents (Price 2008).[17] In modern data, first-born children have higher

[17] Analyzing the behavior of 3,755 mothers in the NLSY79, Buckles and Kolka find that mothers are 15.4% less likely to breastfeed a second-born child than a fourth, 6.6% less likely to take prenatal vitamins in a fourth or higher-

measured IQ (Black, Devereux, and Salvanes 2011), higher noncognitive abilities, and a higher likelihood of working in occupations with leadership skills (Black, Gronqvist and Ocker 2018).

Consistent with these findings, scientists in the MoS have fewer siblings and are more likely to be the first-born child (Table 2, Panel B): The average scientist who was a child in 1880 had 2.6 siblings, and the average star had 2.7 siblings, significantly less than the 3.2 average for boys of the same age in the population. Similarly, 34.1% of scientists and stars each were first-born compared with 27.9% of boys in the population.

Controlling for family size and birth order, however, leaves the estimate for *High SES* substantially unchanged, at 0.382 (significant at 1%, Table 6, column 6), implying a 46.5% change in the odds of becoming a star, nearly identical to the baseline odds of 46.9%. Estimates for the *number of siblings* and for being a *first-born* are negative but not statistically significant, with p-values of 0.243 and 0.472, respectively.

Using servants as a measure of SES slightly increases the estimated class gap in recognition. Controlling for family size and birth order, scientists who grew up with servants have 69.9% higher odds of becoming stars (Table A2, column 6), up from 66.7% in the baseline.

Are the Children of Immigrants Less Likely to be Stars?

While high-skilled immigration has encouraged US innovation (e.g., Moser, Voena, and Waldinger 2014), many immigrants to the United States have been low-skilled, drawn to the United States by a heightened demand for unskilled workers (Rosenbloom 2002) or pushed from their homeland by poverty (e.g., O'Rourke 1997, pp.775-76). When these immigrants arrived in the United States, they had to overcome hardships that made it difficult for them and their children to succeed (e.g., Collins and Zimran 2023), and possibly become academic stars.

We find that a person's nativity status is much less important than their SES. Controlling for *foreign-born* parents leaves the estimate for *high-SES* at 0.384 (Table 6, column 7 and Figure 5, Panel A), implying a 46.9% change in odds. Moreover, the estimate for *foreign-born* parents is close to zero (at 0.026, with a p-value of 0.884). Equivalent estimates for servants confirm that *foreign-born* parents cannot explain the outsized influence of SES (with an estimate of 0.511,

order birth than in a first and are 10.6 % less likely to receive early prenatal care. Using data from the American Time Use Survey, Price (2008) finds that first-born children receive 20-30 additional minutes of quality time per day compared with second-born children in the same family.

implying 66.7% higher odds, Table A2, column 7, and Figure 5, Panel B).

Are the Children of Uneducated Mothers Less Likely to be Stars?

Finally, we investigate the effects of mothers, and more specifically their literacy, as a measure of education.[18] Investigating the link between parental education and IQ for 1,528 California school children who scored in the top 1% of the national IQ distribution in 1921, Leibowitz (1974) found that mothers' education is a significant predictor of boys' IQ, while fathers' education is insignificant.[19] Examining the influence of parental education in modern data, Black et al. (2005) find that mothers' education has a positive impact on the education of children in Norway, while there is no effect of fathers' education.

Among scientists in the MoS (1921), having grown up with an illiterate mother reduces the odds of stardom by 67.5% (with an estimate of -1.123, significant at 10%, Table 6, column 8), confirming the important role mothers play even among this selective group of professionals.

Yet, maternal literacy explains only a small share of the correlation between SES and stardom. Controlling for illiterate mothers, high-SES scientists have 45.5% higher odd of becoming a star (just 1.4% less than the baseline of 46.9%) and scientists who grew up with servants have 66.0% higher odds of becoming a star (just 0.7% less than the baseline of 66.7%).

Do Network Connections with Existing Stars Facilitate Stardom?

In their analysis of gender bias in recognition, Card et al. (2022) find that connections with an existing Fellow of the Econometric Society increase one's probability of being nominated and selected into the Society in subsequent years. To evaluate such connections as a mechanism for becoming a star in the MoS, we match scientists with their patents and use co-inventor connections with an existing star as a measure for connections. Charles Parsons, for example, is *connected* with Archibald Campbell (a star in 1910) because they collaborated on patent US656208A for a "reversing steam-turbine." Consistent with existing results on differences in inventors' reliance on patents across industries (Moser 2012), chemistry and physics have the

[18] Just 6% of married mothers of scientists in the MoS worked outside the home; this low rate of employment prevents us from using mothers' occupation as a measure for childhood SES.

[19] These data, which were originally collected by Lewis M. Terman, include 857 boys and 671 girls, with a mean IQ of 151.5 and 150.4, respectively (Leibowitz 1974, p.436). Notably these children tended to taller and stronger than their classmates. Their fathers had a median 12.4 years of schooling and their mothers a median of 11.7 years, nearly 4 years above the U.S. average of their generation.

largest number of scientist inventors: 342 and 294 scientists in chemistry and physics have at least one patent, respectively.[20] Across all 12 disciplines, 489 scientists (including 175 chemists and 136 physicists) have at least one patent with co-inventors.

Star chemists are twice as likely to be connected with an existing star than other scientists: 6 of 29 stars with co-inventors (20.7%) are connected with at least one scientist who had become a star already in 1906 or 1910. By comparison, just 10.3% of other scientists with co-inventors (15 of 146) are connected with an existing star. Adding to the importance of connections, another 10 of the connected non-stars in 1921 became stars in later editions.

To investigate the influence of connections more systematically across all disciplines, we re-estimate equation 5 with indicators for scientists with *patents* and for scientists who are *connected* by a joint patent with a scientist who had been voted a star already. These estimates show that scientists who are connected with an existing star have 208.0% higher odds of becoming a star themselves (Table 6, column 9). Importantly, however, the link between SES and stardom is robust to controlling for patents and connections, which suggests that SES matters even conditional on connections. Controlling for patents and connections, scientists from high-SES homes have 47.7% higher odds of becoming stars (Table 6, column 9), and scientists from households with servants have 88.9% higher odds (Table A2, column 9, Figure 5).[21]

Do Scientists Learn from their Fathers Occupations?

Investigating the influence of early exposure to innovation, Bell et al. (2019) show that the children of inventors are more likely to work in the same field (USPTO technology class) as their fathers. Investigating political dynasties, Dal Bó et al. (2009 and 2017) document intergenerational persistence in politics, where powerful politicians manage to transfer power to family members. To investigate this channel in science, we match the disciplines of stars and other scientists with the occupations of their fathers.

We find strong evidence for intergenerational persistence in medicine. 13.6%, 13.3%, and 15.1% of scientists in pathology, physiology, and anatomy, respectively, were the children of

[20] We observe 1,161 chemists in the MoS (1921) as children. 342 of them (47.8%) have at least one patent before 1921; all of them together created a total of 2,836 patents. 175 of the chemists with a patent have a co-inventor; together, they produce a total of 612 patents with co-inventors. Chemists with co-inventors include 29 stars and 146 non-star scientists. 29 stars have 94 patents with co-inventors, and 146 non-stars have 518 patents with co-inventors.
[21] The estimate for connections stays positive but loses statistical significance when we control for publications and citations in addition to patents, with an estimate of 0.649 and a p-value of 0.154.

physicians, compared with just 3.58% of scientists in other disciplines (Figure A6, Panel A) and 0.7% of boys of the same age in the 1880 census (Figure 2, Panel A). Moreover, 12.9%, 7.8%, and 13.3% of stars in pathology, physiology, and anatomy, respectively, were the children of physicians (Figure A6, Panel B), compared with just 3.63% of stars in other disciplines.

Similarly, having a professor father increases a person's odds of becoming a scientist and becoming a star: 1.7% of scientists and 2.4% of stars are the children of professors; by comparison, just 0.02% of boys in the census of 1880 are the children of professors. Data from a 2017-20 survey of 7,204 U.S.-based tenure-track faculty suggest that nearly one quarter of faculty today have at least one PhD parent, compared with 0.9% of the population and 11.8% of PhD recipients (Morgan et al. 2022, p. 1626). Our results indicate intergenerational persistence at an even more elite level, above the PhD, among scientists and stars.

Moreover, we find that the children of farmers are most likely to become scientists and stars in botany. 51.1% of scientists in botany are the children of farmers, compared with 28.4% of scientists in other disciplines. 42.5% of stars botanists are the children of farmers, compared with 23.8% of stars in other disciplines. Taken together, our results indicate that scientists acquire valuable occupation-specific human capital through their fathers and that such capital increases their odds of becoming scientists—and stars—in a related discipline.

Are the Sons of Wealthy Farmers More Likely to be Stars?

Next, we investigate the influence of SES (and more specifically, parental wealth) within the occupation of "farmers," who account for a large share of our scientists and the population. 29.3% of star scientists whom we observe as children in the census of 1870 were the children of farmers, compared with 31.1% of all scientists and 45.6% of boys of the same age in the population (Figure 3). In the census of 1880, 29.0% of star scientists were farmers, compared with 34.4% of scientists and 55.6% of the population (Figure 2), and in the census of 1900, 14.3% of star scientists were farmers, compared with 30.3% of scientists and 56.2% of the population (Figure 4). Data on personal wealth is available only for the 1870 census. In that year, personal wealth for farmers ranged between \$0 and \$999,997, equivalent to \$23 million in 2023 (measured as the relative share in GDP per capita, MeasuringWorth.com 2024).

Confirming the importance of SES, the sons of farmers with less wealth are underrepresented among scientists and stars (Figure A7). Among the children of farmers, 8.9% of star scientists and 6.9% of all scientists grew up on farms with zero wealth, compared with 12.2% of the population. In 1870, the median star scientists lived in a farm household with $1,400 of personal wealth, compared with $1,000 for all scientists and $500 in the population. Similarly, the average star scientist lived in a farm household with $7,805 of personal wealth, compared with $5,575 for all scientists and $2,788 in the population.

Scientists from the top of the wealth distribution had dramatically higher odds of becoming stars. A scientist whose father was a farmer with more than $20,000 of personal wealth (the top 2.5% of personal wealth for this sample) had 60.5% higher odds of becoming a star than a farmer with similar publications but less than $20,000 in wealth (Table A4, column 2).

V. CONCLUSIONS

To examine the influence of inequality on science, we have applied machine-learning methods to link comprehensive biographical data on scientists with individual census records. Our analyses of these linked records indicate that patterns of underrepresentation that affect innovation today were already present more than a century ago, in the early 1920s. Children of attorneys, physicians, and clergy are overrepresented in science, while children of farmers and other low-income occupations are underrepresented. Using the occupational income rank of a person's father as a measure of childhood SES, we find that the median scientist is drawn from the 78th percentile of SES. Capturing differences in parental education and occupational prestige, alternative measures of SES imply even larger differences in participation.

Beyond participation, we show that SES influences peer recognition in science. Scientists from high OCCSCORE families have 46.9% higher odds of becoming a star, and scientists from families with servants have 66.7% higher odds, controlling for the number and the quality of publications. Mathematics is the only field in which the influence of SES operates through publications; in all other fields, controlling for publications fails to reduce the influence of SES. Using job titles to identify academic scientists, we show that there is a significant prestige penalty of holding industry jobs, and that academia is more egalitarian than industry. Controlling for academic jobs, as well as publications, scientists from high-SES families have 46.9% higher odds of being stars, and scientists from households with servants have 66.7% higher odds. In addition to publications and academic employment, the correlation between SES and stardom is robust to controlling for patents, connections with existing stars, elite graduate degrees, family size, birth order, and maternal education; none of these controls significantly reduces the correlation between stardom and SES.

Consistent with recent findings on the importance of colleges as a mechanism of upward mobility (e.g., Chetty 2017, 2023; Michelman 2023), scientists with elite undergraduate degrees have 41.5% higher odds of becoming stars, controlling for publications. Elite college degrees explain more of the correlation between SES and stardom than any other control; yet, even controlling for elite degrees, high-SES scientists have 43.3% higher odds of becoming stars, and those from families with servants have 58.1% higher odds. These findings emphasize the importance of SES and indicate that a person's class may unduly influence how we perceive their work.

References

Abele, Andrea E., Naomi Ellemers, Susan T. Fiske., Alex Koch, and Vincent Yzerbyt, 2021. "Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups." *Psychological Review, 128*(2), 290–314.

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen, 2018. "The Social Origins and IQ of Inventors." *NBER Working Paper* #24110.

Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva, 2016. "Taxation and the International Mobility of Inventors." *American Economic Review* 106, no. 10: 2930–81.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez. Stefanie Stantcheva, Valerio Sterzi, 2018. "Dancing with the Stars: Innovation through Interactions, *NBER Working Paper* #24466.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman, 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Economic Journal: Applied Economics*, 11(3): 338–381

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang, 2010. "Superstar Extinction." *The Quarterly Journal of Economics* 125, no. 2, 549–589.

Biasi, Barbara, David Deming, and Petra Moser, 2022. "Education and Innovation." in *The Role of Innovation and Entrepreneurship in Economic Growth,* edited by Aaron Chatterji, Josh Lerner, Scott Stern, and Michael J. Andrews, pp.537-575. Chicago: University of Chicago Press,

Biasi, Barbara, Michael Dahl, and Petra Moser, 2024. "The role of bipolar disorder and family wealth in choosing creative occupations. *Nature: Scientific Reports* 14, 10703. https://doi.org/10.1038/s41598-024-61320-y

Biasi, Barbara and Song Ma, 2023. "The Education-Innovation Gap" *NBER Working Paper* #29853.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova and John Van Reenen, 2019. "Who becomes an inventor in America? The importance of exposure to innovation." *The Quarterly Journal of Economics*, *134*(2), pp.647-713.

Black, Sandra E. and Paul J. Deveraux, 2011. “Recent Developments in Intergenerational Mobility.” *Handbook of Labor Economics*, Vol. 4, Part B, pp. 1487-1471.
Black, Sandra E., Erik Grönqvist, and Björn Öckert, 2018. “Born to Lead? The Effect of Birth Order on Noncognitive Abilities.” *The Review of Economics and Statistics*; 100 (2): pp.274–286.
Black, Sandra E., Paul Devereux and Kjell Salvanes, 2005. “Why the Apple Doesn’t Fall Far: Understanding Intergenerational Transmission of Human Capital,” *American Economic Review*, 95, 437–449.
______, 2011. “Older and Wiser? Birth Order and IQ of Young Men.”, *CESifo Economic Studies*, Volume 57, Issue 1.
Buckles, Kasey, and Shawna Kolka, 2014. “Prenatal investments, breastfeeding, and birth order.” *Social science & medicine (1982),* 118, pp.66–70.
Bradley, Robert H., and Robert F. Corwyn, 2002. “Socioeconomic status and child development.” *Annual Review of Psychology, 53*(1), 371–399.
Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri, 2022. “Gender Differences in Peer Recognition by Economists.” *Econometrica,* 90 (5): 1937–1971.
______, 2023. “Gender Gaps at the Academies” *Proceedings of the National Academy of Sciences*. 120 (4)
Cattell, James McKeen ed., 1906. “*American men of science: A biographical directory*." Bowker.
______, 1906. “A statistical study of American men of science.” *Science,* vol. xxiv, no. 622, pp. 699-707
Cattell, James McKeen and Dean R. Brimhall, 1921. “*American men of science: A biographical directory*." Garrison, New York, The Science Press.
Chen, Jiafeng and Jonathan Roth, 2024. “Logs with Zeros? Some Problems and Solutions.” *The Quarterly Journal of Economics*, 139(2), pp. 891–936,
Chetty, Raj, David J. Demming, and John N. Friedman, 2023. “Diversifying Society’s Leaders? The Determinants and Casual Effects of Admission to Highly Selective Private Colleges.” *NBER Working Paper #31492*.
Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, 2017. “Mobility Report Cards: The Role of Colleges in Intergenerational Mobility.” *NBER Working Paper* #23618.
______, 2020. “Income Segregation and Intergenerational Mobility Across Colleges in the United States.” *Quarterly Journal of Economics,*135(3), pp. 1567–1633.
Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, 2014. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *The Quarterly Journal of Economics,* 129(4), pp. 1553–1623.
Collins, William J., and Ariell Zimran, 2023. "Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration." *American Economic Journal: Applied Economics* 15, no. 3, pp.238-269.
Dal Bó, Ernesto, Fred Finan, Olle Folke, Torsten Persson and Johanna Rickne, 2017. “Who Becomes a Politician?” *Quarterly Journal of Economics*, 132(4), pp.1877-1914.
Dal Bó, Ernesto, Pedro Dal Bó, and Jason Snyder, 2009. “Political dynasties.” *The Review of Economic Studies*, *76*(1), pp.115-142.
Duffy, Thomas P., 2011. The Flexner Report - 100 years later. *Yale Journal of Biological Medicine*. (3):269-76. PMID: 21966046; PMCID: PMC3178858.

Einiö, Elias, Josh Feng, Xavier Jaravel, 2024. "Social Push and the Direction of Innovation" Working paper.
Feigenbaum, James, 2016. "Automated census record linking: A machine learning approach."
Freeman, Richard, Danxia Xie, Hanzhe Zhang, and Hanzhang Zhou, 2024. "Institutional Concentration of Elite Researchers". Working paper.
Jetten, Jolanda, Frank Mols, Nikita Healy, and Russel Spears, 2017. "Fear of falling: Economic instability enhances collective angst among societies' wealthy class." *Journal of Social Issues, 73*(1), 61–79.
Heiserman, Nicholas, and Brent Simpson, 2017. "Higher inequality increases the gap in the perceived merit of the rich and poor." *Social Psychology Quarterly, 80*(3), 243–253.
Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, 2019. "The Allocation of Talent and US Economic Growth." *Econometrica* 87 (5): 1439–74.
Integrated Public Use Microdata Series (IPUMS). "Integrated Occupation and Industry Codes and Occupational Standing Variables in the IPUMs." Accessed July 22, 2024. https://usa.ipums.org/usa/chapter4/chapter4.shtml#occscore.
Leibowitz, Arleen, 1974. "Home Investments in Children," *Economics of the Family: Marriage, Children, and Human Capital*, ed. by T. W. Schultz, University of Chicago Press, pp. 432–456.
Michelman, Valerie, Joseph Price, and Seth D Zimmerman, 2022. "Old Boys' Clubs and Upward Mobility Among the Educational Elite", *The Quarterly Journal of Economics*, 137(2), pp. 845–909.
Morgan, Alison C., Nicholas LaBerge, Daniel B. Larremore, Mirta Galesic, Jennie E. Brand, and Aaron Clauset, 2022. "Socioeconomic roots of academic faculty." *Nature Human Behavior* 6: 1625–1633.
Moretti, Enrico, and Daniel J. Wilson, 2017. "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists." *American Economic Review*, 107 (7): 1858–1903.
Moser, Petra, 2012. "Innovation without Patents: Evidence from World's Fairs." *The Journal of Law & Economics*, *55*(1), 43–74.
Moser, Petra, Alessandra Voena, and Fabian Waldinger, 2014. "German Jewish Émigrés and US Invention." *American Economic Review*, 104 (10): 3222–55**.**
Moya, Miguel, and Susan T. Fiske, 2017. "The social psychology of the Great Recession and social class divides." *Journal of Social Issues, 73*(1), 8–22.
National Academy of Sciences, 1978. *The National Academy of Sciences: The First Hundred Years 1863–1963.* Washington (DC): National Academies Press (United States).
National Research Council, 2009. Mathematics Learning in Early Childhood: Paths towards Excellence and Equity. Washington, DC: The National Academies Press. https://doi.org/10.17226/12519.
O'Rourke, Kevin H., 1997. "The European Grain Invasion, 1870-1913." *Journal of Economic History*, 57 (4), pp. 775-801
Pennington, Jeffrey, Richard Socher, and Christopher D. Manning, 2014. "GloVe: Global vectors for word representation." In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543.
Price, Joseph, 2008. "Parent-child quality time: does birth order matter?" *Journal of Human Resources.* 43 (1), pp.240-265.

Rosenbloom, Joshua L., 2002. *Looking for Work, Searching for Workers: Labor Markets during American Industrialization*. Cambridge University Press.

Rossiter, Margaret W., 1982. “Women Scientists in America. Struggles and Strategies to 1940.” The John Hopkins University Press, Baltimore and London.

San, Shmuel, 2023. “Who works were and why? The role of social connections in the labor market. Working paper.

Siegel, Paul M., 1971. “Prestige in the American Occupational Structure.” Unpublished doctoral dissertation. University of Chicago.

Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June Hsu and Kuansan Wang, 2015. “An Overview of Microsoft Academic Service (MAS) and Applications.” In *Proceedings of the 24th International Conference on World Wide Web* (pp. 243-246).

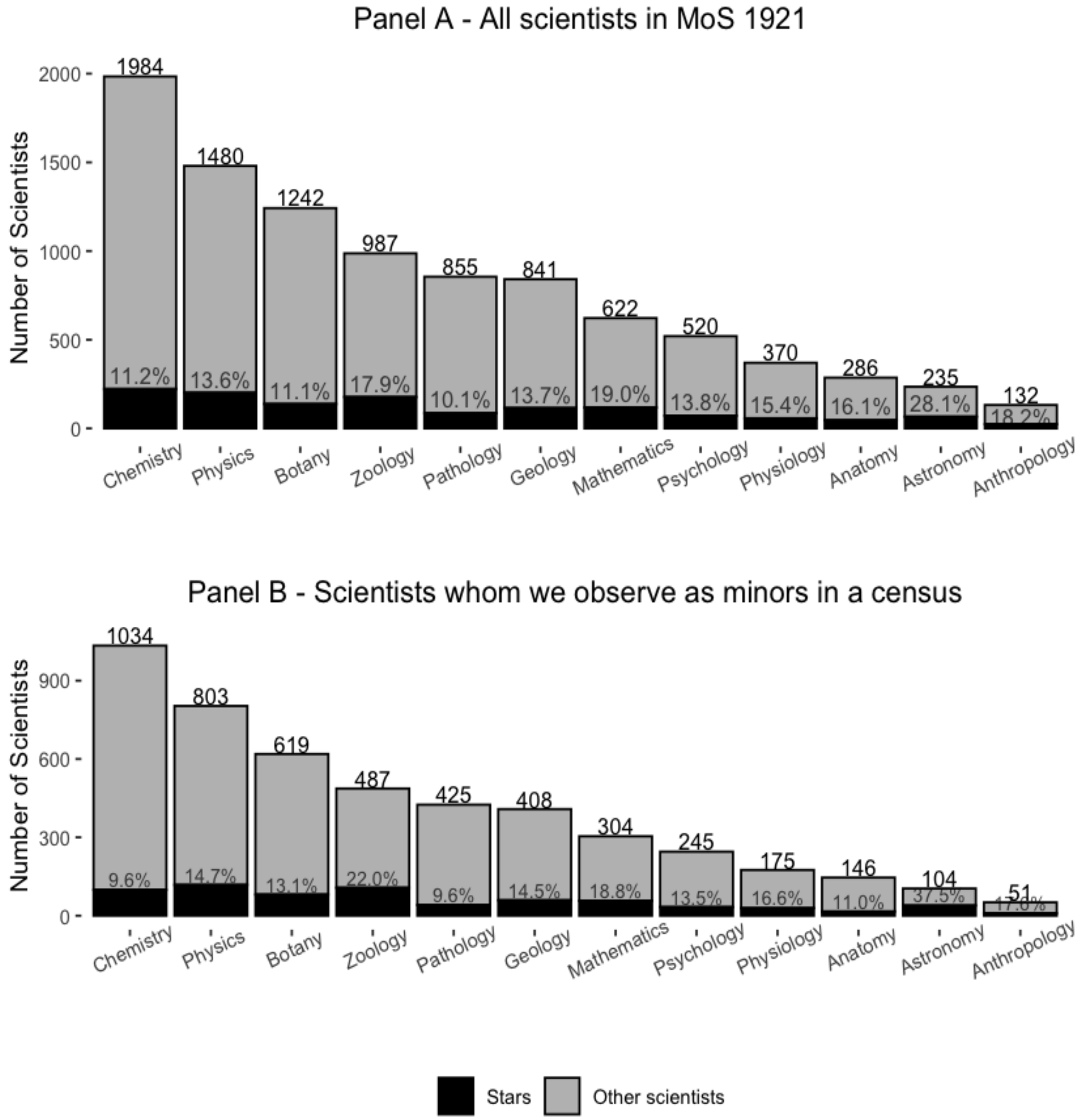
Sobek, Matthew, 1995. "The Comparability of Occupations and the Generation of Income Scores." Historical Methods 28(1): 47-51.

Stansbury, Anna, and Robert Schultz, 2023. “The Economics Profession’s Socioeconomic Diversity Problem.” *Journal of Economic Perspective.* Volume 37, Number 4, pp. 207–230.

Tanjitpiyanond, Porntida, Jolanda Jetten, and Kim Peters, 2022. “How economic inequality shapes social class stereotyping.” *Journal of Experimental Social Psychology, 98,* Article 104248.

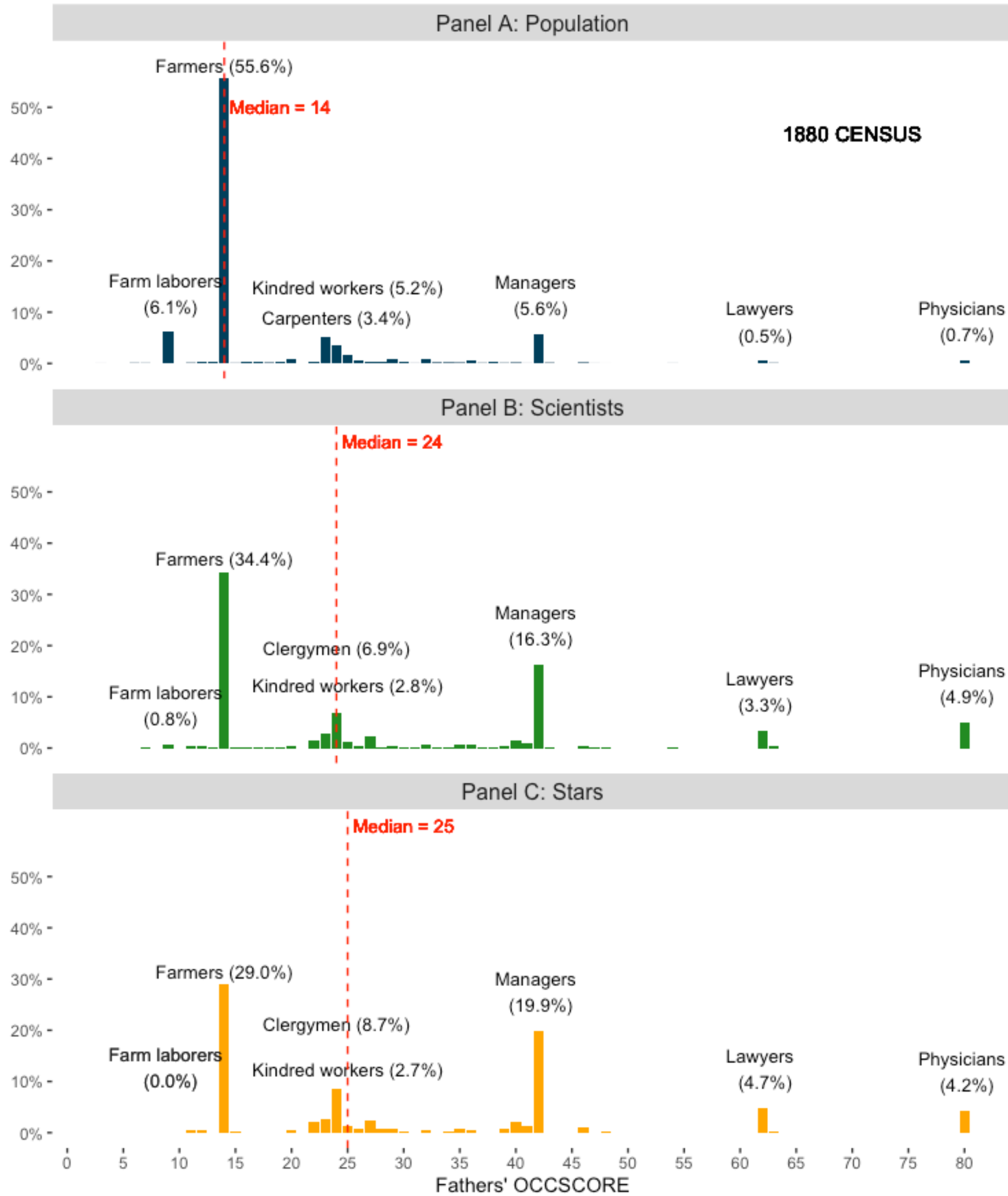
Visher, Stephen Sargent, 1947. *“*Scientists starred, 1903-1943, in American Men of Science.*”* Johns Hopkins Press.

FIGURE 1 – STARS AND OTHER SCIENTISTS ACROSS 12 DISCIPLINES



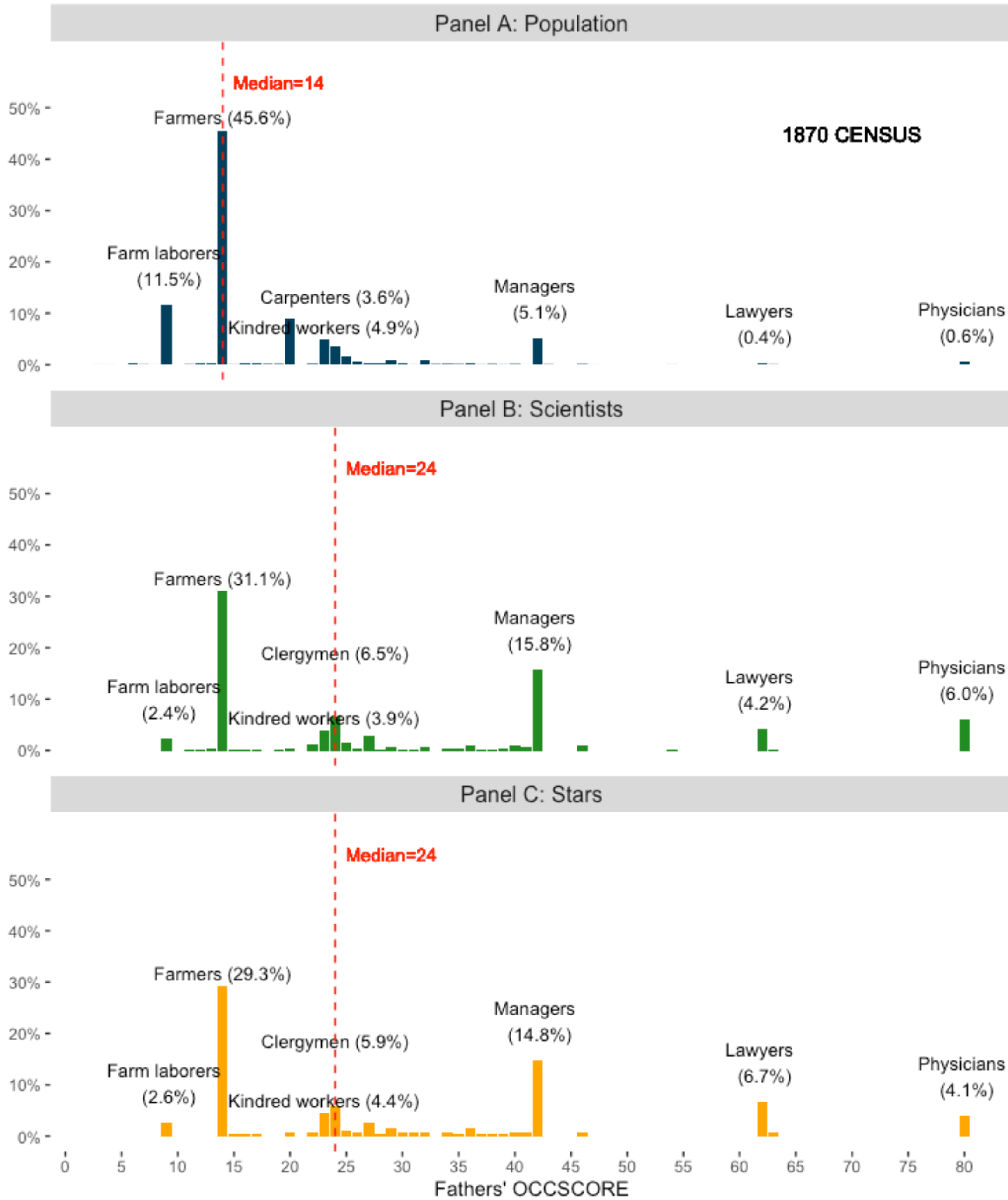
Notes: This figure plots the distribution of scientists across the 12 disciplines, in which scientists chose stars among their peers. Disciplines are arranged by their size (measured by the number of scientists in that discipline, in grey). The number on top of each bar represent the total number of scientists in each discipline. The black shaded area represents the number of stars; we report the share of stars above the black bar. Disciplines are observable for stars in Visher (1947), but not for other scientists. To assign the remaining (non-star) scientists to disciplines, we use the text that describes the research of stars to create training data for a nearest-neighbor algorithm. This algorithm assigns each of the remaining scientists uniquely to one of the 12 disciplines. Panel A includes all 9,554 scientists, Panel B includes 4,067 scientists whom we can match with at least one census wave when they are minors. These 4,067 scientists, which form the data set for our main analyses, are distributed similarly across disciplines to the full sample (Appendix Figure C1).

FIGURE 2 –SES (MEASURED BY FATHERS' OCCSCORE) FOR SCIENTISTS AND STARS COMPARED WITH BOYS OF THE SAME AGE IN THE CENSUS OF 1880



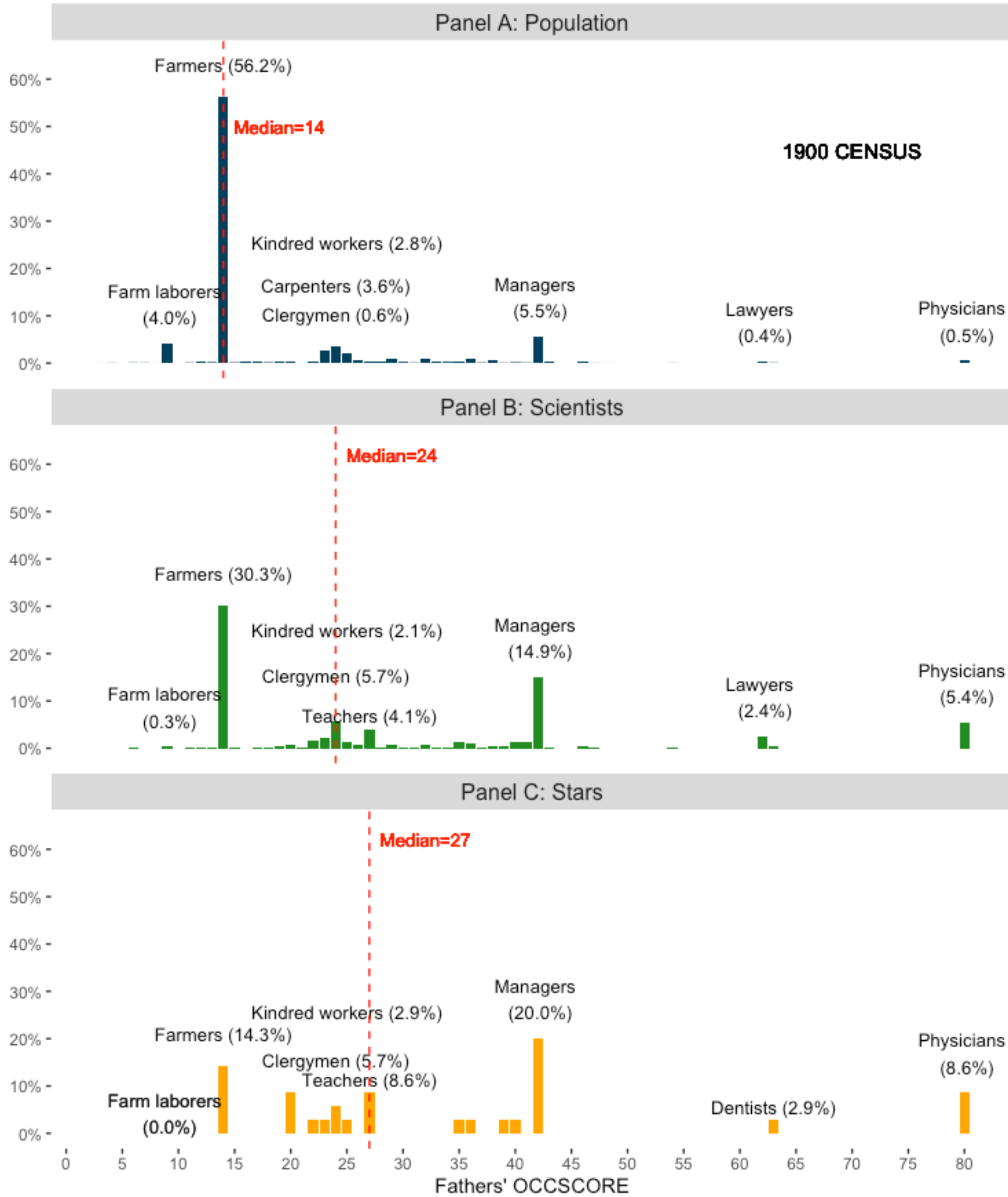
Notes: To investigate whether stars and other scientists are disproportionately drawn from high-income families, we plot the distribution of fathers' OCCSCORE for scientists (Panel B) and stars (Panel C) against the distribution of boys of the same age in the population (Panel A). Data include 2,274 scientists whom we observe as minors in the census of 1880.

FIGURE 3 – SES (MEASURED BY FATHERS' OCCSCORE) FOR SCIENTISTS AND STARS COMPARED WITH BOYS OF THE SAME AGE IN THE CENSUS OF 1870



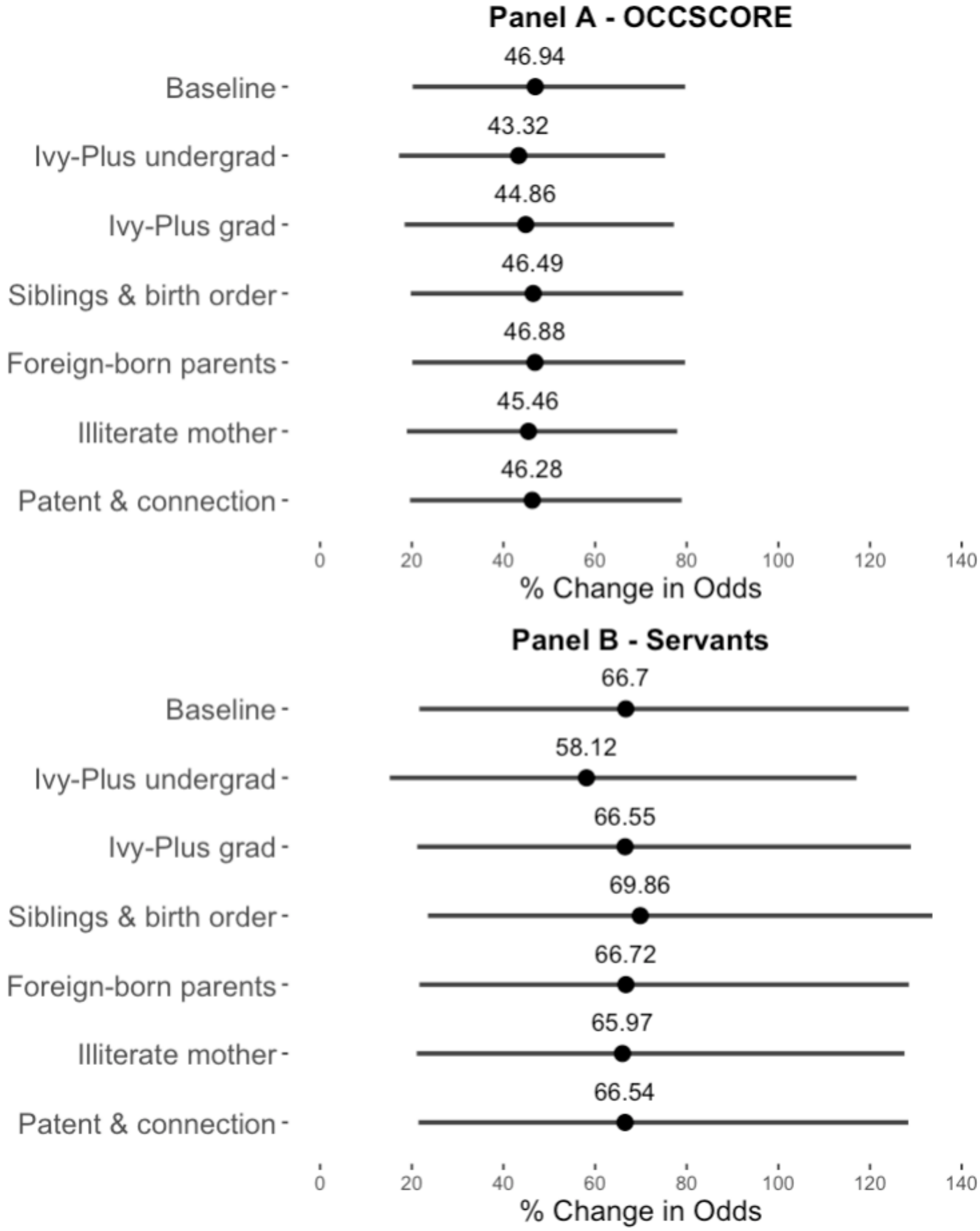
Notes: To investigate whether stars and other scientists are disproportionately drawn from high-income families, we plot the distribution of fathers' OCCSCORE for scientists (Panel B) and stars (Panel C) against the distribution of boys of the same age in the population (Panel A). Data include 1,142 scientists whom we observe as minors in the census of 1870.

FIGURE 4 – SES (MEASURED BY FATHERS' OCCSCORE) FOR SCIENTISTS AND STARS COMPARED WITH BOYS OF THE SAME AGE IN THE CENSUS OF 1900



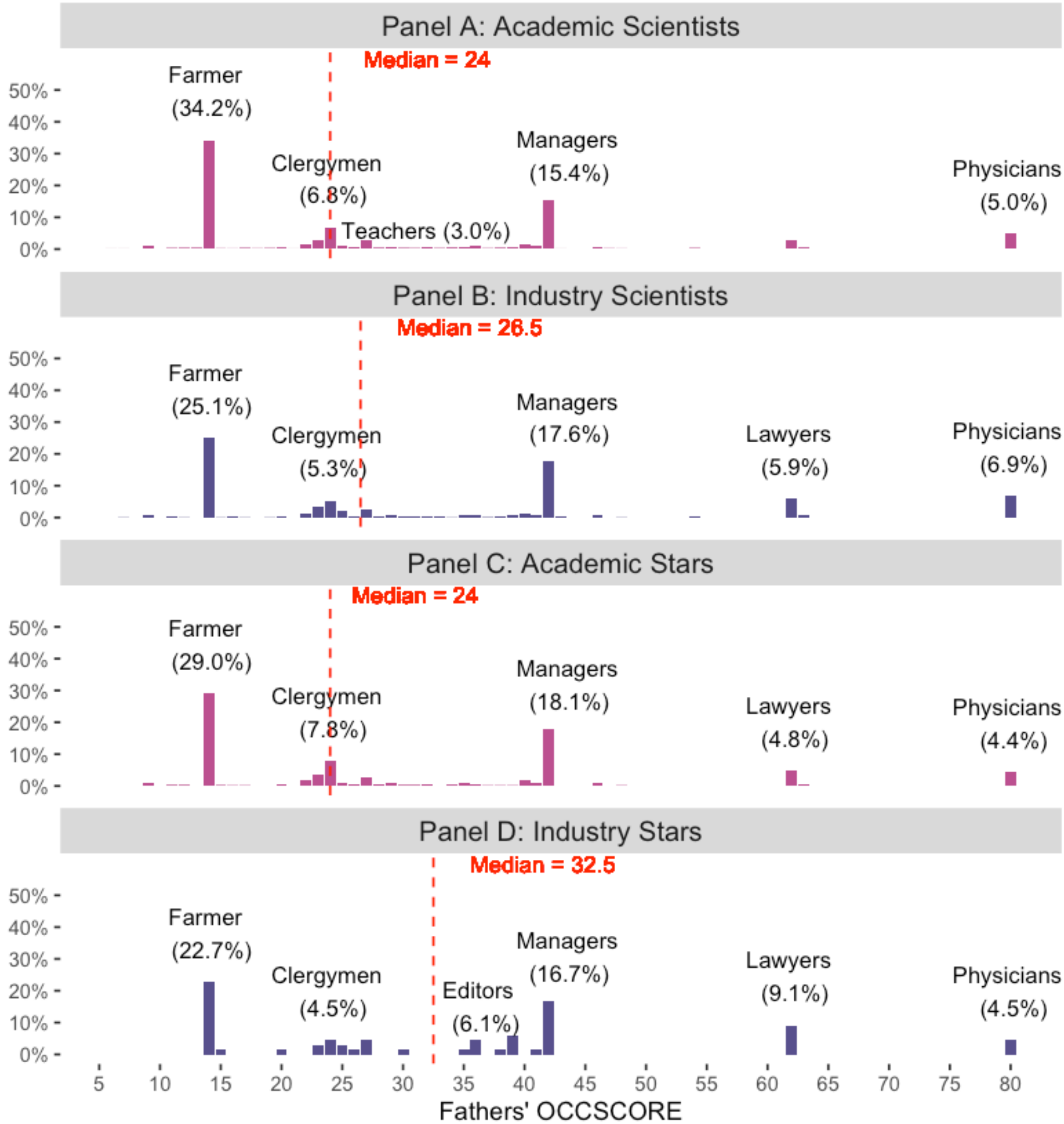
Notes: To investigate whether stars and other scientists are disproportionately drawn from high-income families, we plot the distribution of fathers' OCCSCORE for scientists (Panel B) and stars (Panel C) against the distribution of boys of the same age in the population (Panel A). Data include 1,182 scientists whom we observe as minors in the census of 1900.

FIGURE 5 – CONTROLLING FOR ELITE EDUCATION AND FAMILY TRAITS



Notes: To investigate alternative mechanisms for the link between SES and stardom, we re-estimate equation 3 with controls for *Ivy-Plus undergraduate* and *graduate* degrees, *siblings* and being *first-born*, having *foreign-born parents* or an *illiterate mother*, as well as having a *patent* and being *connected* by a joint patent with a star. Baseline estimate includes control for industry jobs, publication and citations (Table 6 and Table A2, column 3). Panel A plots the *% Change in Odds* (reported in Table 6) with 95% confidence interval for scientists from high-OCCSCORE families and panel B plots the *% Change in Odds* (reported in Table A2) with 95% confidence interval for scientists from families with servants.

FIGURE 6 – THE CHILDHOOD SES OF SCIENTISTS AND STARS IN ACADEMIA VS. INDUSTRY



Notes: To investigate whether academic and industry scientists are disproportionately drawn from high-SES families, we plot the distribution of fathers' OCCSCORE for academic scientists (Panel A) against the distribution of fathers' OCCSCORE for industry scientists (Panel B) for the pooled sample of scientists whom we match with at least one census wave in 1870, 1880, or 1900. Panels C and D plot the same comparison for stars. Academic scientists are those who held an academic position, such as a lecturer, assistant professor, or associate professor, at least once in their career. Data include 3,694 academic scientists, 830 industry scientists, 642 stars in academia, and 66 stars in industry.

TABLE 1 – MATCHING SCIENTISTS WITH THEIR CHILDHOOD HOME IN THE CENSUS

	US-born	US-born male	Matched with the US census		Living with their father	
	N	N	N	Rate (in %)	N	Rate (in %)
			Panel A: All scientists			
1870	2,753	2,658	1,484	55.83	1,176	79.25
1880	5,209	5,009	3,189	63.67	2,493	78.17
1900	8,146	7,791	4,687	60.16	1,724	36.78
Any census	8,146	7,791	6,104	78.35		
			Panel B: Scientists who were minors in a census year			
1870	2,409	2,319	1,334	57.52	1,142	85.61
1880	4,083	3,910	2,566	65.63	2,274	88.62
1900	2,693	2,550	1,644	64.47	1,182	71.90
Any census	7,549	7,215	4,866	67.44	4,067	83.58

Notes: Counts and shares (in %) of scientists linked with the US census. Panel A covers the full sample of US-born, male scientists, irrespective of their age in a census year. Panel B focuses on scientists who were minors in a census year; we match these scientists with individual records in the US census. "Living with their father" includes US-born male scientists whose father's occupation we can observe in the census. *Any census* reports data for scientists we match with at least one census wave. For 531 scientists we observe father's occupation in both 1870 and 1880; we use the earlier census, reducing the number of observations for the 1880 census to 1,743 and creating a regression sample of 4,067 scientists (*N* in Panel B, column *Living with their father*, and row *Any census*). Individual census records for 1890 were destroyed by a fire in 1921.

Table 2 –Traits of Scientists and their Childhood Homes Compared with the Population

	1870 Census		1880 Census		1900 Census	
	Scientists	Population	Scientists	Population	Scientists	Population
N	1,334	427,892	2,566	539,416	1,644	787,878
Panel A: Childhood SES Measured by Father's Occupation (Median)						
OCCSCORE	24.0	14.0	24.0	14.0	24.0	14.0
ERSCORE	44.1	9.9	44.1	9.9	56.7	9.9
EDSCORE	6.3	4.6	16.4	4.6	17.1	4.6
Duncan SEI	32.0	14.0	44.0	14.0	47.0	14.0
SIEGEL	40.7	40.7	40.7	40.7	40.7	40.7
Panel B: Other Traits of the Scientist's Childhood Home (Mean and SD)						
Servant			0.086	0.004	0.241	0.014
			(0.01)	(0.00)	(0.02)	(0.00)
Siblings	2.304	2.979	2.571	3.234	2.202	3.364
	(0.05)	(0.00)	(0.04)	(0.00)	(0.05)	(0.00)
First-born	0.400	0.280	0.341	0.279	0.507	0.324
	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Foreign-born parents	0.068	0.058	0.088	0.078	0.106	0.093
	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Illiterate mother	0.017	0.176	0.007	0.159	0.005	0.096
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: This table compares scientists whom we observe in their childhood home with a 5% stratified sample of children of the same age in the same census year. Panel A reports the median of alternative measures of SES: OCCSCORE and ERSCORE for fathers' occupational income ranks, EDSCORE for fathers' education, Duncan SEI and SIEGEL for occupational prestige. Panel B reports means (and SD) for other traits. *Servants* indicates scientists who grew up in households with at least one live-in servant; this variable is available for 1880 and 1900 but not 1870. *Siblings* counts the number of siblings living in the same household and *first-born* indicates scientists who are the oldest child in their household. The variable *foreign-born parents* indicates scientists who have at least one foreign-born parent. *An illiterate mother* can neither read nor write.

TABLE 3 –LOGIT ESTIMATES FOR THE ODDS OF BEING A STAR AS A FUNCTION OF CHILDHOOD SES

	(1) OCCSCORE	(2) EDSCORE	(3) ERSCORE	(4) Duncan SEI	(5) SIEGEL	(6) Servant
High SES	0.327***	0.407***	0.300***	0.368***	0.425***	0.564***
	(0.095)	(0.095)	(0.096)	(0.095)	(0.095)	(0.146)
Age	0.044***	0.044***	0.043***	0.045***	0.043***	0.072***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Constant	-4.223***	-4.269***	-4.173***	-4.290***	-4.238***	-5.555***
	(0.490)	(0.490)	(0.490)	(0.490)	(0.490)	(0.517)
% Δ in odds	38.66%	50.28%	35.04%	44.43%	52.98%	75.71%
% Δ in probabilities	33.46%	43.37%	30.41%	38.41%	45.59%	63.90%
Discipline and census FE	Y	Y	Y	Y	Y	Y
N	4,067	4,067	4,067	4,067	4,067	4,067

Notes: Logit estimates of $P(star) = \left(1 + e^{-(\alpha+\beta SES_{high\,SES}+\delta_c+\delta_d+\gamma age)}\right)^{-1}$ where $SES_{high\,SES}$ indicates scientists whose father had an occupation in the top half of SES when the scientist was between 0 and 18 years old. OCCSCORE and ERSCORE measures the occupational income rank of the father's occupation. EDSCORE measures the share of people in that occupation who have attended college. Duncan and Siegel capture the prestige of the father's occupation. *Servant* indicates scientists who grew up in families with at least one live-in servant. We calculate *% Δ in odds of being star* as $e^{\hat{\beta}} - 1$ and *% Δ in probability of being star* as $\frac{P(star|SES_{high\,SES}=1)}{P(star|SES_{high\,SES}=0)} - 1$. For more details on these calculations see Appendix D.

TABLE 4 –LOGIT ESTIMATES OF THE ODDS OF BEING A STAR CONTROLLING FOR PUBLICATIONS AND CITATIONS

	No controls for pubs.	*asinh*	Top 1%	Top 5%	Top 10%	ln(+0.01)	Pub in *Science*
Panel A: High SES is high OCCSCORE							
High SES	0.327***	0.324***	0.320***	0.335***	0.341***	0.309***	0.303***
	-0.095	-0.102	-0.096	-0.099	-0.1	-0.1	-0.1
Publications		0.079	1.188***	1.163***	0.938***	-0.054	0.381***
		-0.056	-0.332	-0.174	-0.135	-0.044	-0.056
Citations		0.440***	2.286***	1.809***	1.540***	0.305***	0.003
		-0.043	-0.381	-0.186	-0.146	-0.036	-0.009
% *Δ* in odds	38.66%	38.26%	37.71%	39.79%	40.64%	36.21%	35.39%
Panel B: Scientists from households with servants							
	No controls for pubs.	*asinh*	Top 1%	Top 5%	Top 10%	ln(+0.01)	Pub in *Science*
Servants	0.564***	0.443***	0.528***	0.572***	0.504***	0.437***	0.549***
	(0.146)	(0.155)	(0.152)	(0.154)	(0.157)	(0.153)	(0.157)
Publications		0.088	1.239***	1.115***	0.836***	-0.075	1.019***
		(0.060)	(0.384)	(0.188)	(0.147)	(0.048)	(0.156)
Citations		0.088	1.239***	1.115***	0.836***	-0.075	1.019***
		(0.045)	(0.403)	(0.177)	(0.147)	(0.039)	(0.121)
% *Δ* in odds	75.70%	55.78%	69.56%	77.18%	65.61%	54.88%	73.13%
Age, discipline, and census year FE	Y	Y	Y	Y	Y	Y	Y

Notes: To investigate whether scientists are stars because they publish more, we re-estimate the baseline (column 1) with controls for the quantity and quality of publications, in addition to age, discipline, and census year FE. Our preferred specification (in column 2) controls for the inverse hyperbolic sine of a scientists' pre-1921 publications and citations to these publications. Alternative measures for quality include indicators for scientists in the top 1, 5, or 10% of publications and citations within their discipline as well as publications in *Science*. Panel A includes 4,067 scientists whose fathers' occupation we can observe in the census of 1870, 1880, or 1900. Panel B includes 4,132 scientists whom we observe in the census of 1880 and 1900, when census data includes information on servants.

TABLE 5 – LOGIT ESTIMATES – WITHIN DISCIPLINES - OF THE ODDS OF BEING A STAR

	Chemistry (N = 864)		Physics (N = 684)		Botany (N = 538)		Zoology (N = 403)	
High SES	0.214	0.321	-0.063	-0.044	0.640**	0.713**	0.346	0.185
	(0.24)	(0.254)	(0.23)	(0.239)	(0.27)	(0.318)	(0.26)	(0.290)
Publications		0.177		-0.040		0.111		-0.260
		(0.163)		(0.119)		(0.173)		(0.177)
Citations		0.270**		0.422***		0.558***		0.631***
		(0.123)		(0.095)		(0.123)		(0.135)
% Δ in odds	23.89%	37.90%	-6.12%	-4.34%	89.65%	104.01%	41.31%	20.36%
	Pathology (N = 352)		Geology (N = 332)		Mathematics (N = 260)		Psychology (N = 223)	
High SES	0.711*	1.379***	0.775**	1.124***	0.551*	0.316	0.737*	0.856*
	(0.39)	(0.518)	(0.32)	(0.347)	(0.33)	(0.397)	-0.399	(0.449)
Publications		-0.008		0.408		0.326		0.637**
		(0.286)		(0.250)		(0.205)		(0.285)
Citations		0.619***		0.278*		0.508***		0.426**
		(0.178)		(0.166)		(0.134)		(0.172)
% Δ in odds	103.60%	297.09%	117.06%	207.71%	73.50%	37.10%	108.97%	135.37%
Age FE and census FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: To investigate whether the link between childhood SES and stardom varies across disciplines, we estimate equation (1) separately within each of the 12 disciplines, using father's OCCSCORE as a measure of SES. Disciplines are important because they serve as the comparison group within which scientists rank each other. They are, however, only observable for stars and not for other scientists. To address this issue, we use the text that describes the research of stars to train a nearest-neighbor matching algorithm and use this algorithm to assign all scientists uniquely to one of the disciplines. *Publications* and *citations* are transformed using the inverse hyperbolic sine. All estimates include age and census year FE. Robust standard errors in parentheses.

TABLE 6 –LOGIT OF THE ODDS OF STARDOM AS A FUNCTION OF CHILDHOOD SES (OCCSCORE), PUBLICATIONS, AND OTHER TRAITS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
High SES	0.327***	0.394***	0.385***	0.360***	0.371***	0.382***	0.384***	0.375***	0.390***
	(0.095)	(0.096)	(0.103)	(0.103)	(0.103)	(0.103)	(0.103)	(0.103)	(0.097)
Industry jobs		-1.096***	-1.164***	-1.137***	-1.072***	-1.166***	-1.164***	-1.174***	-1.119***
		(0.155)	(0.178)	(0.177)	(0.179)	(0.177)	(0.178)	(0.178)	(0.158)
Publications			0.113**	0.116**	0.119**	0.113**	0.113**	0.113**	
			(0.057)	(0.058)	(0.058)	(0.058)	(0.057)	(0.058)	
Citations			0.423***	0.417***	0.413***	0.423***	0.423***	0.426***	
			(0.042)	(0.043)	(0.043)	(0.042)	(0.042)	(0.043)	
Elite undergrad				0.347***					
				(0.110)					
Elite grad					0.398***				
					(0.105)				
N siblings						-0.038			
						(0.033)			
First-born						-0.085			
						(0.119)			
Foreign-born parents							0.026		
							(0.177)		
Illiterate mother								-1.123*	
								(0.575)	
Patent									-0.029
									(0.115)
Connect									1.125***
									(0.403)
% Δ in odds	38.66%	48.26%	46.94%	43.32%	44.86%	46.49%	46.88%	45.46%	47.69%

Notes: Industry job indicates scientists who work in industry, rather than academia before 1921. *Patent* is an indicator for scientists who patented at least 1 invention by 1921. *Connect* indicates scientists who have patented with at least one star before 1921. Elite undergrad and elite grad indicate scientists who have completed a degree at an Ivy Plus university (as in Chetty et al. 2019). *First-born* scientists indicates scientists who are the oldest child in their household. *Foreign-born parents* indicates scientists with at least one foreign-born parent. An *illiterate mother* can neither read nor write. All estimates include age, discipline, and census year FE; robust standard errors in parentheses.