THE MS. ALLOCATION OF TALENT

MORITZ LUBCZYK, ROCKWOOL FOUNDATION BERLIN, AND
PETRA MOSER, NYU, NBER, AND CEPR*

JULY 09, 2024

We investigate whether changes in the allocation of talent could close the innovation gender gap. Analyzing the biographies of 70,000 scientists – linked with their patents – we find that the underrepresentation of women in innovation today is a continuation of a long-run trend that was already in place for scientists born in the 1920s. OLS estimates indicate that the low share of female scientists in patent-intensive STEM fields is the main driver of this innovation gender gap. To identify the causal effects of gender differences in the allocation of talent, we exploit an exogenous shock in female representation due to WWII, when the scarcity of male scientists pulled female scientists into patent-intensive research fields in STEM . Using variation in enlistment as an instrument for female entry, we find that one additional woman becomes an inventor for every five women entering STEM. Counterfactual estimates imply that, if women were as likely to work in STEM fields as men, the innovation gender gap would close in 37, rather than 118 years. Using a three-stage Roy (1951) model to interpret these results, we show that women face distortions in entering science, entering STEM, and in patenting. Due to these distortions, women have to be substantially more talented to enter science.

KEYWORDS: INNOVATION, SCIENCE, GENDER, AND MISALLOCATION

*We thank Christian Dustmann, Martin Fiszbein, Marita Freimane, Jacob French, Jacopo Gambato, and seminar participants at Barnard, Bayreuth, LSE, Northwestern, Nottingham, NYU, Católica, Rockwool Foundation Berlin, Rockefeller University, and Zurich for helpful comments and discussions.

As talented women and Black men have moved into high-skilled professions, improvements in the allocation of talent account for 20-40% of productivity growth since the 1960s (Hsieh, Hurst, Jones, and Klenow 2019). Yet, women continue to be underrepresented in science and invention, creating a persistent innovation gender gap. In 2018, just 32% of US PhDs in the physical sciences were women (European Commission 2021). For every nine US patents filed by an all-male team between 1985 and 2019, there was just one patent with at least one female inventor (OECD 2023). Fewer than 20% of inventors born in 1980 were women; at current rates of convergence, it will take 118 years to achieve gender parity (Bell, Chetty, Jaravel, Petkova, and Van Reenen 2019).

Can changes in the allocation of talent close the innovation gender gap? To investigate this question, we analyze rich biographical data on 70,039 male and female scientists from the *American Men of Science* (MoS 1956), a comprehensive compendium on US scientists. To analyze changes in the innovation gender gap, we first link scientists with their patents. These data show that the slow speed of convergence today (Bell et al. 2019) is a continuation of a trend that was present already nearly a century ago (Figure 1).

To investigate the drivers of this persistent gender gap in innovation, we estimate OLS regressions of the probability that a scientist becomes an inventor on an indicator for female gender and control variables for productivity (publications), holding an industry job (versus working in academia), holding a PhD, marriage and children, and doing research in the social, biological, or physical sciences. In these regressions, a scientist's research focus explains most of the gender differences in patenting. Simply put, women work in fields in which male scientists patent less: 36.8% of scientists in the physical sciences are inventors, but just 3.3% of physical scientists are women. The share of women is much higher in the social sciences (9%), but patenting in the social sciences is rare.

To identify the causal effects of changes in the allocation of talent we exploit an exogenous shift in the allocation of talent due to WWII, when the scarcity of male scientists pulled female scientists into male-dominated, patent-intensive research fields. By matching scientists with the *U.S. World War II Army Enlistment Records, 1938-1946*, we create a measure for exposure to enlistment across fields. To assign scientists to fields, we apply a k-means clustering algorithm (with k=100) to the text that describes the research of each scientist. In the median field, 19% of draft-eligible male scientists enlisted in WWII. Difference-in-differences regressions indicate that fields with an additional 10-percent increase in the share of enlisted male scientists experienced a 41% increase in female entry. Event study estimates show that female entry into fields with high enlistment rates peaks with a 92% increase in 1942 and remains high until 1950. Fields with an above median enlistment rate experience a 40% increase in female entry. Using enlistment as an instrument for female entry, we find that, for every five women entering a field in the physical sciences, one additional woman becomes an inventor. In a complementary set of IV regressions at the

individual level, we instrument the fields of female entrants in year t with the leave-one-out field-shares of other female entrants in the same year. Individual-level estimates imply that entering the physical, rather than social sciences, increases a female scientist's propensity to patent by 11.6%.

We use these estimates to construct reweighting counterfactuals (DiNardo, Fortin, and Lemieux 1996) that illustrate the impact of talent allocation on gender parity. Counterfactual estimates imply that 61% more women born between 1940 and 1980 would have become inventors if women were as likely to enter STEM as men; these additional women would have produced up to 2.6% more patents per birth year cohort. If women entered STEM at the same rate as men, gender parity in innovation would be reached in 37 years - 81 years earlier than the 118 years given current rates of convergence (Bell et al. 2019).

We interpret these findings through the lens of three-stage Roy (1951) model of entry into science, STEM fields, and patenting. Using this model, we predict the share of women at each stage assuming 1) that men and women draw from the same distribution of talent and 2) that women face no distortions. Comparing the predictions of this model with the observed data reveals major distortions at each stage. Examining differences in implied talent thresholds for women and men, we find that women must be much more talented to enter science than men. Since only the most talented women become scientists, reduced competition and positive selection increase wages for both women and men in science, encouraging some men below the talent threshold to enter science. Using data on PhD degrees and elite education to investigate selection, we find that female scientists are positively selected. Importantly, selection is weaker during WWII, when the shortage of male labor temporarily reduces entry barriers for women.

These findings contribute evidence on misallocation of talent – by gender – to the literature on inequality in innovation. Hsieh et al. (2019) show that between 20 and 40% of growth in aggregate market output per person between 1960 and 2010 can be explained by improvements in the allocation of talent as women and Black men have entered skilled occupations. Focusing on gender inequality among university scientists, however, Iaria, Schwarz, and Waldinger (2022), document persistent gender gaps in hiring, publishing, and promotions for women, even though the share of women at elite universities has increased from just 2% in 1900 to nearly 18% in 2000. Linking patents with tax records, Bell et al. (2019) show that children from below-median income families are ten times less likely to become inventors than children in the top percentile of parental income, and that 82% of 40-year old inventors today are men. We extend this literature by providing causal estimates of changes in the allocation of talent on female innovation and by quantifying the distortions that lead to female underrepresentation in science and innovation.

Our findings also contribute to understanding the effects of WWII on the US economy. Much of this literature has focused on the war's effect on female labor supply (e.g. Goldin

1991; Acemoglu, Autor, and Lyle 2004; Goldin and Olivetti 2013; Jaworski 2014; Rose 2018). Combining census data with retrospective surveys on married women's labor force participation between 1940 and 1950, Goldin (1991) shows that up to 54% of women who entered the labor force during the war had left again by 1950. To identify the causal effects of the war on female labor force participation, existing research has used geographical variation in enlistment rates across states (Acemoglu et al. 2004; Fernández, Fogli, and Olivetti 2004; Goldin and Olivetti 2013) and within states over time (Jaworski 2014).[1] A separate strand of research has examined the war's long-run effects on innovation, e.g., through changes in public spending on research (Gross and Sampat 2023), or changes in intellectual property rights (Biasi and Moser 2021). We connect these literatures by investigating the effects of changes in female labor force participation on innovation.

1. Historical Background

Before WWII, female scientists were relegated primarily to "women's work." In her study of *Women Scientists in America* Margaret Rossiter (1982, p. 51) writes: "If by 1910, women had succeeded in being allowed to earn degrees from almost all German and American universities granting them to men, they were far less successful in these years in gaining equal treatment in the world of employment. The resistance to women's holding the same jobs as men was much stronger than it was to women's earning the same degrees as men, largely because most jobs were traditionally labelled or 'sex typed.' [...] a separate labor market for women emerged in the sciences in the 1880s and 1890s, when they began to seem scientific employment in significant numbers, and it was firmly established in several fields by 1910. This practice of 'sex segregration' was usually justified with the essentially conservative rhetoric that women had 'special skills' or "'unique talents' for certain fields or kinds of work."

Countless women failed to find any employment as scientists; 138 of them are listed in the MoS (1938) as "unemployed." Others chose to work in more traditional women's occupations - "schoolteaching, editorial work (for scientific journals, abstract journals, the Science Service, and the *American Men of Science* itself), translating, writing, librarianship, and office work as administrative assistants and executive secretaries (Rossiter 1982, pp. 262-3). For instance, Gertrude B. Elion – 1988 Nobel laureate – worked as a high school teacher in New York until 1942.

[1] Acemoglu et al. (2004) show that states with a 10% higher mobilization rate experienced a 9.8% increase in women's labor force participation between 1940 and 1950. Fernández et al. (2004) show that US states with a 10% higher mobilization rate experienced a 29% increase in female labor force participation for individuals whose mothers worked during the war. Studying the labor force participation of married white women in 1950 and 1960, Goldin and Olivetti (2013) show that the persistent effects of WWII on labor force participation were driven by married women with 12 or more years of schooling and without children. Jaworski (2014) shows that, in states with median enlistment rates, women obtained 0.16 fewer years of schooling, reducing their competitiveness after the war. Rose (2018) shows that wartime investments production were a key driver of increased in female labor force participation during WWII but that most of these effects had faded by the 1950s.

Holding women to high standards, university faculties "were almost adamantly opposed to advancing or promoting any but the most extraordinary" (Rossiter 1982, p. xvi). One instance that crystallized this attitude was a publicity campaign surrounding Marie Curie's visits to the United States in 1921 and 1929, followed by a biography in 1938 and a wartime movie. By allowing "claims to a position in science to be linked (however illogically) to Curie's exceptional career, [Curie's followers] probably thought that they were implying that any American women could follow in her footsteps. But before long most professors and department chairmen were interpreting Curie's example far more restrictively and expecting that every female aspirant for a faculty position must be a budding Marie Curie. They routinely compared American women scientists of all ages to Curie and finding them wanting, justified not hiring them on the unreasonable grounds that they were not as good as she, twice a Nobel Laureate! [...] The impact of Curie's visit thus came to mean less an 'opening of doors' for women in sicence [...] and more [...] a 'raising of the thresholds' to almost unattainable heights" (Rossiter 1982, p.127).

Limits on the participation of women in science became impractical after September 1940, when the Selective Training and Service Act initiated the United States military draft. The basis for the draft were draft registrations of eligible men; local draft boards assessed registrants for their capability to serve (Acemoglu et al. 2004). Initially, registration requirements covered all men between 21 and 35 years of age and included a number of deferments for dependent family members (wives and children), war-relevant occupations (in particular, farm work and food supply), and ability to serve (physically and mental ability, but also race, and German, Italian, or Asian ancestry). As the war progressed, registration requirements rapidly extended to cover all men aged 18-44 and, by 1944, the draft deferments were almost entirely eliminated (Goldin and Olivetti 2013). As effectively all men of eligible age registered for the draft, the main driver of variation in enlistments across locations (Acemoglu et al. 2004) or, in our case, research fields, was the rate of drafting among eligible men.[2]

2. DATA

Our main data cover the lives and careers of 70,039 male and female scientists in the MoS (1956) linked with their patents, publications, and enlistment records to identify scientists who served in WWII.

[2]Table 4 in Acemoglu et al. (2004) provides a quantitative assessment of variation in state-level enlistment rates using census data. The authors find significantly lower enlistment rates in states with higher shares of employment in farming and German, Italian, or Japanese ancestry. In turn, enlistment was significantly higher in states with higher average schooling among the local population and in states with a larger population share aged 13-24 in 1940 (a 13-year-old in 1940 would be 18 and eligible in 1945).

Biographies of American Scientists

To collect biographical data, we have digitized the text of all 82,094 unique biographies in the *American Men of Science* (MoS 1956).[3] Despite its name, the MoS includes male and female scientists in Canada and the United States. A comprehensive data base of scientists, these data were originally collected by James McKeen Cattell, the first professor of psychology in the United States and the first editor of *Science* for nearly 50 years. (Cattell 1906, p.v) collected the MoS for his own research, creating "for the first time a fairly complete survey of the scientific activity of a country in a given period." With support from the Carnegie Institution, Cattell published his data for the "chief service [. . .] to make men of science acquainted with one another and with one another's work." He continued to update the MoS until his death; the 1956 edition was published by his son Jacques.

Entries were based on membership in scientific societies (such as the American Mathematical Society or the American Society of Bacteriologists) and subject to a comprehensive review from "scientific societies, universities, colleges, and industrial laboratories." In the editor's preface, Cattell (1956) thanks "thousands of scientific men who have contributed names and information about those working in science," and "acknowledges the willing counsel of a special joint committee of the American Association for the Advancement of Science and the National Academy of Science National Research Council "which acted in an advisory capacity."

Detailed career data allow us to distinguish scientists working in academia and industry. We use job titles, such as professor, research fellow, or instructor, to identify 52,785 scientists, who worked in academia at least once. Our data include 3,525 female academic scientists (87.7% of all female scientists in the MoS) and 49,260 male academic scientists (74.6% of all male scientists). In our data, 17,254 scientists work exclusively in industry; 495 of these scientists are women and 16,759 are men. We use data on the year when a scientist started their first job identify the career start year for all scientists. Men are 25.4 years old on average when they start their careers in science (with a median of 25); women are slightly older at 25.9 years old (with a median of 24).

Female Scientists, Birth Years, Birthplaces, Marriages, and Children

To identify female scientists, we use historical gender frequencies of first names in the US Social Security Administration Records (SSA) between 1880 and 2011, implemented in Python's *gender-detector* package. This approach outperforms hand-matching in a comparison benchmarked on graduates of women's colleges as true positives (Kim and Moser 2024). The algorithm assigns the gender of 70,780 scientists (86.2% of the total 82,094), including 4,220 women and 66,560 men.

[3]This count excludes 6,352 duplicate entries who appear in more than one of the three volumes of the MoS (1956), as well as 2,549 scientists whose entry consists only of a reference to another edition of the MoS.

We use data on birth years to investigate changes in productivity across the life cycle, and to control for age and cohort fixed effects. We also exploit birth years to create a high-quality match between scientists and their patents. Birth years are available for 81,461 scientists (99.2%). We know both the birth years and gender for 70,230 scientists (85.5%).

Data on each scientist's place of birth allow us to capture the influence of a scientist's childhood environment on female innovation. For instance, (Bell et al. 2019, Table V, column 2) show that women who grew up in a neighborhood with one standard deviation more female inventors are 21.3% more likely to patent. We control for such variation through birthplace fixed effects at the level of US states. We observe birth places for 81,693 scientists (99.5%), 70,661 of them are US-born (86.5%, 2,751 were born in Canada (3.4%), and 8,281 (10.1%) in other countries.

A key advantage of the MoS is that it records the scientist's year of marriage and children. For example, Yolanda Tota Pratt married in 1943 and had one child by 1956:

PRATT, DR. YOLANDA T(OTA), College of Arts & Sciences, University of Maryland, College Park, Md. ORGANIC CHEMISTRY. Elmira, N. Y, July 18, 13; m. 43; c. 1. A.B, Cornell, 38; Ph.D.(chem), Columbia, 42. Lecturer chem, Barnard Col, Columbia, 41-42; asst. Harvard, 42-43; ORG. RESEARCH CHEMIST, Hercules Powder Co, Wilmington, 43-45; MARYLAND, 47- With nat. insts. health, U. S. Pub. Health Service. Chem. Soc. Pyrazines; resin acids; quinolinequinones.

Assigning Scientists to Research Fields

We investigate gender differences in the allocation of talent across 3 "sciences" and 100 research fields. The MoS (1956) separates scientists into the physical sciences (35,339 individuals, 3.3% women), the biological sciences (21,740, 7.7% women), and the social sciences (12,969, 9.0% women). To investigate gender differences in the allocation of talent at a more fine-grained level, we apply a *k*-means matching algorithm (with k=100) to the text that describes' each scientist's research topics and their discipline. Research topics are known for 96.4% of 82,094 scientists in the MoS (1956); disciplines are reported for 90.2% of scientists. Dr. Pratt, for example, describes her work as "Pyranzines; resin acids; quinolinequinones."

Intuitively, *k*-means clustering works like a multi-dimensional least square algorithm, which groups together data points (here, scientists) that are most similar in terms of their observable characteristics (here, research topics). A "cluster" (here a field) refers to a collection of data points (scientists) that are grouped together because they have similar traits (here topics). To group scientists into clusters, the *k*-means algorithm assigns researchers to one of the k centroids by minimizing the distance between the observations and the centroid.

A major advantage of using *k*-means is that it captures meaningful connections across scientists' research. For instance, Eugene E. Howe and Charles Tanford list their disciplines

as pharmaceuticals and chemistry, respectively. Both work on amino acids: Howe examines "synthesis of amino acids; isolation of amino acids; parenternal amino acids; ion exchange processes . . . " and Tanford research interests include "polarography; metal complex formation of amino acids; peptides and proteins; physical chemistry of proteins". The k-means algorithm captures this overlap and assigns both scientists to a field in the area of biochemistry.

We combine fields and the sciences in a nested structure by assigning each field to the science of the median scientist. For example, Howe's and Tanford's field in "biochemistry" is assigned to the physical sciences because 85.7% of scientists are in the physical sciences.

Matching Scientists with Patents

To measure the gender innovation gap and capture changes in innovation by American scientists across research fields, we count the number of successful patent applications by each scientist per field and year. We construct these data through an improved matching process that links all 82,094 American scientists with their US patents between 1910 and 1970 from Google Patents, using information on the scientist's full name, their age, and their discipline.[4] To assess the match quality between scientists and patents, we use patents that are matched to scientists when they are between 0-18 years old as a measure for false positive matches, and choose matching rules that minimize the share of false positives. Using full names (including middle name), and excluding the 20% most frequent names, we reduce the error rate from 83.3% with the most naïve matching without middle names to 6.3% and 4.2% in the physical sciences.

Name-based patent matching may overstate the innovation gender gap if we miss a significant share of patents by women who change their name upon marriage. In practice, however, we are unlikely to miss many patents because just one third of few female scientists married and entries for married scientists typically report their maiden names.[5]

Comparing patenting rates for male and female scientists, we find that female scientsts are significantly less likely to patent: 3.6% of female scientists have at least 1 patent compared with 23.2% of male scientists (Table 1).

Matching Scientists with WWII Enlistment Records

To measure variation in exposure to enlistments across fields, we link scientists with archival enlistment records from the US National Archives. Each record includes the enlisted soldier's name (e.g., "Ellis V. Jr. Hunt"), their Army serial number ("170076256"), year of enlistment

[4]Moser, Parsa, and San (2024) use these data to investigate the effects of nationality-based restrictions on immigration in the 1920s on US innovation.

[5]Just 38.8% of female scientists in the MoS (1956) married (1,560 total scientists), and just 466 married women worked in the physical sciences, where scientists are most likely to patent. 434 (93.1%) of these scientists report their middle name, which for married women is typically the maiden name. We use this information match them with the census and the patent record.

("1942"), their branch (INF, for "infantry"), "nativity" (or birthplace, "Missouri), education ("4 years of high school"), and marital status ("single, without dependents.")

Enlistment records include 9,025,234 men and women enlisted in the US Army and Air Force between 1938-46. They cover all branches of the Army, including soldiers in combat roles (e.g., the Air Corps), service support (e.g., the Quartermaster Corps, responsible for food and clothing), and communications (e.g., Signal Corps, responsible for the maintenance of radar technologies). Enlisted also include personnel in the Corps of Engineers Chemical Warfare Service and in physical health.[6]

Matching 8.4 million unique soldier names with 82,904 unique scientist names, yields more than 500 billion potential matches. To address this issue, a fast heuristic algorithm jointly applies N-Grams, Term Frequency - Inverse Document Frequency (TF-IDF), and Cosine Similarity algorithms to identify likely matches between soldiers and scientists based on full names with middle names. A second set of algorithms addresses the issue that the records of the National Archives are scanned with an error rate of 5%; these algorithms use phonetic matching (Double Metaphone, Philips 2000) based on primary and secondary pronunciations with tokenized distance algorithms (Mean Square Contingency or "Pearson II" 1913; Iterative Substring, Stoilos, Stamou, and Kollias 2005; Levenshtein et al. 1966). In a third step, we use scientist's birth year, birthplace, education, and enlistment years to filter out improbable matches. Finally, we use information on military service in the MoS (1956) to fine tune the matching parameters. By this process, we match 10,285 scientists in our sample with a unique enlistment record.[7] Counts of enlisted scientists increase during the war and decline in 1945 (Figure 2), mirroring changes in US government spending and industrial mobilization (e.g., Rose 2018). 39% of male scientists born between 1918 and 1927 enlisted in the army between 1941 and 1946, 19% of male scientists in the median field enlist (Figure 3).

Matching Scientists with Publications in Microsoft Academic Graph

To measure individual-level differences in research productivity, we match scientists with their publications in Microsoft Academic Graph (MAG, Sinha, Shen, Song, Ma, Eide, Hsu, and Wang 2015).[8] We match scientists in the MoS to their MAG author identifier using their

[6]Since the Air Force was part of the Army until 1947, Air Force records are included in enlistments under "US Army Air Corps" until 1941. Enlistments for the US Navy, however, are not recorded in the National Archives. If these scientists were exposed to combat, including them in the control would attenuate our estimates. 1,855 scientists report service in the Navy; we exclude them from the main regressions.

[7]For additional detail on the matching process, see Chen and Moser (2024), who developed the matching to investigate the long-run effects of combat on academic scientists; provide additional details on the matching process. We use a broad definition of draft eligibility to include 54,424 men between 18 and 44 during WWII.

[8]Moser and Parsa (2024) use these data to investigate the effects of political persecution during the McCarthy era on American science. 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 (1956) with *authorids* in the MAG, using first and last names, as well as middle initials. For scientists who are matched with more than one author, we manually check and remove duplicates.

first, last, and middle names. To minimize false positives, we focus on English-language publications and authors with at least one publication between 1900 and 1970. We limit publications to books and journal articles, and duplicate matches are manually removed.

On average, scientists produced 16.6 publications between 1900 and 1970. With 980 articles and books, Albert Sabin, the inventor of the oral polio vaccine, has the largest number of publications. The endocrinologist Miriam Elizabeth Simpson, holder of the first PhD in anatomy conferred from the University of California, is the most published female scientist, with 249 publications.

3. Female and Male Scientists with Patents

Why are women underrepresented among inventors? To investigate this question, we start by analyzing gender gaps in patenting for a population of highly skilled scientists who have the talents and technical expertise required to patent. Decomposing the gender gap in innovation, we ask what share of the gap is due to gender differences in productivity, academic vs. industry employment, marriage, children, and research fields.

OLS: Decomposing the Innovation Gender Gap

To investigate the contributions of alternative potential drivers of the innovation gender gap, we estimate linear probability models:

$$\mathbb{1}[Inventor]_i = \alpha + \beta_1 \mathbb{1}[Female]_i + \beta_2 X_i + \beta_3 X_i \times \mathbb{1}[Female]_i + \varepsilon_i \tag{1}$$

where the indicator variable $\mathbb{1}[Inventor]_i$ equals 1 if scientist i files at least one successful patent application over the course of their career and $\mathbb{1}[Female]_i$ is an indicator for female scientists. The coefficient β_1 estimates the innovation gender gap. To evaluate the explanatory power of different forces, we observe the change in β_1 conditional on subsets of explanatory variables in X_i and the interaction between X_i and $\mathbb{1}[Female]_i$. The matrix X_i includes potential explanatory variables for the gender gap in innovation. Specifically, we control for productivity, measured by *publications*, to account for potential differences in research productivity levels between male and female scientists. The control variable *PhD* is an indicator for scientists who hold a doctoral degree, to measure differences in qualifications. The variable *industry* distinguishes scientists working in industry (rather than in academia, where promotions are typically tied to publications, rather than patents). The variable *married* distinguishes married scientists of both genders. For women, marriage may reduce opportunities to patent, if women cannot realize the best labor market match due to what Mincer (1978) called the "tied migration of working wives"[9] or if they are required to spend a

[9]Mincer (1978, p. 18-19) emphasizes effects on unemployment and labor force participation. "Both unemployment and labor force withdrawals are results of tied migration of working wives. If employment and earning opportunities turn out to be meager or inferior at destination, withdrawal from the labor force is a likely

larger share of their time on unpaid household labor. A control variable for *children* identifies scientists who have least one child. We include this variable because children affect the timing and the quantity of academic output for mothers (as Kim and Moser (2024) have shown for publications). Indicators for the *physical* and *biological sciences* separate scientists working in these sciences from social scientists as the omitted category. Birth year and birth state fixed effects account for variation in patenting across birth cohorts and across states, e.g., if women who grew up in locations with more female inventors are more likely to patent (Bell et al. 2019). Table 1 provides summary statistics across all variables.

OLS Estimates

OLS estimates illustrate the magnitude of the innovation gender gap at baseline: Female scientists are 19.4% less likely to patent than male scientists born in the same year and in the same state (Table 2, column 1). Differences in research productivity explain little of the innovation gender gap. While scientists of either gender are more likely to patent if they publish more, the innovation gender gap remains almost unchanged at 19.2% (Table 2, column 2). Controlling for having a PhD reduces the gender innovation gap to 16.6% (Table 2, column 3). Women are more likely to have a PhD (63.8%) than men (57.7%). Notably, having a PhD increases the probability to patent for male, but not for female scientists. Controlling for industry employment reduces the gender gap to 15.5% (Table 2, column 4). Scientists of either gender are more likely to patent if they work in industry jobs rather than in academia, but women are much less likely to work in industry than men. Just 12.3% of female scientists work in industry compared with 25.4% of men. Moreover, industry employment appears to create a larger boost to patenting for men (+19.3%) than women (+6.1% Table 2, column 4). Controlling for marital status reduces the gender gap to 15.0% (Table 2, column 5); controlling for children leaves it at 15.5% (Table 2, column 6). Notably, married scientists and those with children are more likely to patent if they are male but less likely to patent if they are female. Finally, male scientists more than twice as likely to be married (84.3% compared with 38.8%) and more than three times as likely to have children (74.0% vs. 22.1%).

By far the largest share of the gender gap is explained by differences in the allocation of male and female scientists across fields (Figure 4). Controlling for research fields in the physical sciences reduces the estimated gender gap to 6.7% (Table 2, column 7), and controlling for both the physical and biological sciences reduces the gender gap to 4.3% (Table 2, column 8).

Controlling for all variables jointly (Table 2, column 9), reduces the estimated gender gap to 0.8% (not statistically significant, with a p-value of 0.22). Confirming the link between academic productivity and patenting, scientists of either gender are more likely to patent if

possibility." Tied migration may reduce patenting by reducing labor force participation or by reducing the quality of the labor market match between the female scientist and her job.

they publish more. The estimate, however, is small: the probability of patenting increases by just 0.9% for a one standard deviation increase in the number of publications. Men are 12.1% more likely to patent if they are in industry, and women are 6.1% more likely. Men with a PhD are 6.7% more likely to patent, compared with just 0.7% for women. Fathers are 3.1% more likely to patent than other men, while mothers are 1.2% *less likely* to patent than other women. Men are 29.5% and 3.0% more likely to patent if they are in the physical and biological sciences, rather than the social sciences, respectively. Women are 9.6% and 2.2% more likely to patent if they do research in the physical and biological, rather than women the social sciences.

Thus, OLS estimates point to research fields as the key determinant of the gender gap in innovation. We now extend this analysis to investigating causal effects.

4. Event Studies: Effects of Enlistment on the Allocation of Scientists

We now begin to investigate the causal effects of gender differences in the allocation of scientific talent across research fields. Specifically, we show that shortages of scientists during WWII drew women into fields that were differentially exposed to enlistment.

To investigate whether differences in exposure to enlistment across research fields drew women into science, we compare changes in female entry after 1939 across fields that were differentially exposed to male enlistments. Specifically, we estimate OLS event studies:

$$Y_{kt} = \alpha_k + \phi_t + \sum_{t, t \neq 1939}^{T} \left(\mathbb{1}[T = t] \times D_k \right) \beta_t + \varepsilon_{kt} \tag{2}$$

The outcome variable Y_{kt} counts female scientists entering field k in year t. Field fixed effects α_k control for variation in female entry across fields that is constant over time, e.g., as a result of persistent professional norms that require long and inflexible hours. Year fixed effects ϕ_t control for variation in female entry over time, e.g., due to an increase in female labor force participation that is shared across fields.

The treatment variable D_k measures the share of enlisted male scientists in field k between 1940 and 1946 relative to the number of draft-eligible scientists in field k. The excluded period is 1939, the last full year before enlistments under the Selective Training and Service Act (Pub. L. 76–783, 54 Stat. 885) started in October 1940. Under the assumption that, had WWII not forced male scientists into military service, changes in female entry would have been comparable for fields with different shares of enlisted male scientists, coefficients β_t estimate the causal effect of male enlistment on female entry across fields.

To investigate this identifying assumption, we first compare the pre-1940 characteristics of fields that differ in their exposure to enlistment. One potential challenge to the identifying assumption would be that enlistment rates were high in fields that were more friendly to women before the war. Importantly, fields that were more exposed to enlistment

after 1940 did not experience more entry by women scientists before 1940. (Appendix Figure B1). OLS estimates show that there is no significant correlation between a field's exposure to enlistment after 1940 and the number of women entering that field before 1940 (Appendix Table A.1, column 1, p-value=0.45), the number of female inventors in that field before 1940 (column 2, p-value=0.20), or the size of that field before 1940 (column 3, p-value=0.13). Working in a specific field also fails to predict enlistment for individual draft-eligible male scientists (Appendix Figure B2). We compare the raw means underlying our event study in Appendix Figure B3: Fields with above and below median enlistment rates show similar trends in female entry before the start of the draft. From 1940 on, female entry increases in both groups of fields, but the effect is much larger for fields with above median enlistment. Female entry remains higher in those fields even after the end of the war.

Further supporting the identifying assumption, estimates for β_t are close to zero and not statistically significant before the war. A joint F-test statistic of 0.38 fails to reject the hypothesis that the coefficients are jointly equal to zero with a p-value of 0.77. To identify the average treatment effect of enlistment on female entry across fields we estimate difference-in-differences OLS regressions:

$$Y_{kt} = \alpha_k + \phi_t + \left(\mathbb{1}[t > 1939] \times D_k\right)\gamma + \varepsilon_{kt} \tag{3}$$

where $\mathbb{1}[t > 1939]$ is an indicator for years after the beginning of the WWII draft and D_k measures the share of enlisted male scientists relative to draft-eligible scientists in field k and γ estimates the average treatment effect of an increase in the share of enlisted men.

Event study estimates confirm that female entry increased in response to male enlistment (Figure 5). In fields with an additional 10% increase in male enlistment, female entry increases by 65.4% in 1941, 92.4% in 1942, 68.1% in 1943, and 49.9% in 1944 (significant at 1 percent) and 19.2% in 1945 (not significant, with a p-value=0.33). After the war, women continue to enter fields that were more exposed to enlistments during the war. In fields with a 10% higher share of enlistment during the war, female entry increases by 39.2% in 1946 relative to 1939, 65.3% in 1947, 69.3% in 1948, 83.6% in 1949, and 76.7% in 1950.

Difference-in-differences OLS estimates indicate that fields with a 10% higher share of enlisted scientists experienced a 41.9% increase in female entry (with a 3.861 estimate for β in Table 3, column 2, significant at 1-percent, relative to the mean of 0.922 female entrants per field in 1939). This increase is robust to alternative definitions of treatment, as well as Poisson and inverse hyperbolic sine (IHS) specifications. Fields in which the share of enlisted men exceeds the median share of 19% enlistment rates experience a 37.9% increase in entry (with an estimate of 0.349 in Table 3, column 1). Fields in the top quartile of enlistment experience a 50.6% increase in female entry (column 3). Estimates from a quasi-maximum likelihood Poisson panel regression imply a 35.7% increase in female entry for each additional 10% share in enlisted men (column 4), and IHS indicate a 44.1% increase.

WWII drew both younger women (below the age of 20) and older women (between the age of 28 and 40) into science (Appendix Figure B4). Younger women pursued entry-level jobs as aides to male scientists in industry, while older women, many of them married and with PhDs, filled teaching jobs left behind by men who had joined the war effort (Rossiter 1982). The share of married women among entrants increased from 9% to 19% for the 1942-43 entry cohorts, compared with 1935-38 entrants (significant at 1%). Similarly, the share of women working in industry jobs increased from 14% to 19% (significant at 5%, Appendix Figure B5).

5. Instrumental Variable Estimates Across Fields

In this section, we analyze whether female entry into traditionally male-dominated and patent-intensive research fields increases female innovation in these fields. OLS estimates may underestimate the causal effect of increased female participation on female innovation if women, for instance, were more likely to end up in research positions with a high degree of administrative or teaching responsibilities that prevented them from engaging in patentable research before the enlistment shock. Alternatively, OLS estimates may overestimate the effect of women entering a specific field if women are more likely to enter fields in which female innovation increases for reasons other than an increase in the number of female scientists, e.g., if a shift in gender norms increases both the share of female scientists in a given field and increases patentable research by women.

To identify the causal effects of changes in the allocation on female innovation, we use the enlistment of male scientists as an instrument for female entry. Data on the number of men and women entering science reveal a dramatic decline in men entering science after 1940, followed by an increase in the number of women entering science (Figure 6). Relative to 1939, the last year before the draft, female entry increased by 77.9% in 1942, 62.8% in 1943, and 19.8% in 1944. Male entry, in contrast, increased by just 6.4% in 1942 and then declined by 21.9% in 1943 and 48.1% in 1944.

Difference-in-differences estimates indicate that this change was driven by variation in male enlistment across fields. We use these results to estimate the effect of one additional woman entering a field on the number of female inventors in that field. Specifically, we estimate the first-stage equation:

$$\text{Women}_{kt} = \psi \times \widehat{\text{Women}}_{kt} + \zeta_k + \omega_t + \xi_{kt} \tag{4}$$

where the outcome variable Women_{kt} counts women entering field k in year t. Field fixed effects ζ_k capture time invariant differences in female entry across fields, for instance, due to persistent norms and inflexible working conditions. Time fixed effects ω_t control for changes in female entry over time. The instrument $\widehat{\text{Women}}_{kt}$ is the predicted share of women pulled

into field k conditional on the long-term trend of women entering that field:

$$\widehat{\text{Women}}_{kt} = \widehat{\text{pre-war trend}}_{k|t<1940} + \gamma \times \text{enlistment}_k \times \mathbb{1}[t > 1939] \tag{5}$$

The parameter γ is the estimate (from equation 3) of the causal effect of exposure to enlistment on female entry. The treatment variable enlistment_k is the share of male scientists in field k enlisting in the army, and $\mathbb{1}[t > 1939]$ indicates years after the draft. The product of a field's enlistment rate, the estimated proportional impact on female entry, and the post-1939 dummy captures the size and timing of the enlistment shock in field k. $\widehat{\text{pre-war trend}}_{k|t<1940}$ is a field-specific linear pre-war trend in female entry. To illustrate the intuition behind our IV specifications, we plot the average number of female scientists (Figure 7) and female inventors (Figure 8) entering the physical sciences over time. After the WWII draft, more women enter fields with higher rates of enlistment among male scientists. In the physical sciences, this increase in the number of female scientists results in an increase in the number of female inventors. Figure 9 shows that the relationship between male enlistment and female entry holds across all fields: correlating the natural log of the number of men enlisted in the army with the natural log of the number of female entrants in the same field after the start of the draft in 1940 results in a correlation coefficient of $\rho = 0.61$.

Using 2SLS, we estimate the first stage jointly with the second-stage equation:

$$Y_{kt} = \mu \times \text{Women}_{kt} + \eta_k + \theta_t + \varepsilon_{kt} \tag{6}$$

The outcome variable Y_{kt} is the number of female scientists-inventors - scientists who create at least one patented invention in their career - entering field k in year t. The endogenous variable Women_{kt} is instrumented by $\widehat{\text{Women}}_{kt}$ in the first-stage equation (4). Field fixed effects η_k control for time-invariant differences in female patenting across fields, and year fixed effects θ_t control for changes over time.

Under the exclusion restriction that the enlistment shock affects the number of female scientist-inventors entering field k in year t only by increasing the number of female scientists entering field k in year t, the coefficient μ estimates the causal impact of an additional female scientist joining field k on the number of female inventors. We estimate single instrument 2SLS equations separately for the physical and biological sciences; μ estimates the causal effect of a woman joining a field on the number of female inventors compared with other fields in the physical or biological sciences.

IV estimates for the physical sciences indicate that, for each woman entering a research field in the physical sciences, the number of female inventors increases by 0.195 (Table 4, column 4); equivalent estimates for the biological sciences indicate that, for each women entering a field in the biological sciences, the number of female inventors increases by 0.049 (Table 5, column 4). Thus, 1 additional woman becomes an inventor for every 5 women

entering the physical sciences and for every 20 women entering the biological sciences.

Notably, the OLS estimate is substantially smaller than the 2SLS estimate both for the physical and biological sciences, suggesting that OLS underestimates the true effect on female innovation. Specifically, the OLS estimate for the physical sciences (0.095) is 51% smaller than the 2SLS coefficient, and the OLS estimate for the biological sciences (0.039) is 21% smaller than the 2SLS estimate.

Downward bias in the OLS estimates suggests that women who enter fields in the physical and biological sciences endogenously are less likely to engage in invention than women who enter "exogenously." Intuitively, women may be more likely to be assigned lower-ranking or administrative tasks that are further removed from cutting-edge research and offer fewer opportunities to patent.

IV estimates also indicate that the field-level enlistment shock is a strong and meaningful instrument for female entry. The predicted number of female entrants is strongly correlated with the actual number of women entering research fields after the beginning of WWII (with Kleibergen-Paap F-statistics of 29 for the physical sciences and 163 for the biological sciences). The magnitudes for the first stage coefficients suggest some variation in compliance across fields. In the physical sciences, 0.175 women join for each woman predicted to join; in the biological sciences 0.536 women join for every woman predicted to join. This difference in first stage strength correlates with the overall representation of women who comprise 3.3% of scientists in the physical sciences and 7.7% in the biological sciences in 1956 (Figure 10). This variation in take-up suggests that the increase in female entry failed to fully compensate for the gap created by male enlistments, possibly due to discrimination (Rossiter 1982, p.29).

6. Scientist-Level IV Estimates

In this section, we analyze the causal effect of entry on female innovation at the level of individual scientists. While field-level analyses identify the causal effect of female entry on innovation across fields, individual-level analyses identify the effects of a person's choice of research field on their probability of becoming an inventor. This distinction is important because, all else equal, scientists may select into research fields for which they have more talent. Analyses at the scientist-level allow us to compare the effects of being in a field conditional on talent, for the same individual.

Specifically, we ask: how much more likely is the same female scientists to become an inventor if she enters a STEM field – the biological or physical sciences – instead of the social sciences? OLS estimates may underestimate this effect if women are predominantly hired for non-research tasks, even within fields with a high propensity to patent. Alternatively, OLS may overestimate the effects of research field choice if female entrants are positively selected on ability and would have become inventors in other fields.

Empirical Strategy: IV Estimates Across Individual Scientists

To estimate the causal effect of entering a specific field on an individual woman's probability to invent, we instrument individual field choices with the field choices of all other women who enter science in the same year. We estimate 2SLS regressions for the population of women in science:

$$\mathbb{1}[Inventor]_i = \beta_{iIV,phy} \times \mathbb{1}[\text{Phys. Sc.}]_i + \beta_{iIV,bio} \times \mathbb{1}[\text{Bio. Sc.}]_i + \theta_t + \varepsilon_i \tag{7}$$

where the outcome variable $\mathbb{1}[Inventor]_i$ indicates female scientists who create at least one patented invention in their career. Indicator variables $\mathbb{1}[\text{Phys. Sc.}]_i$ and $\mathbb{1}[\text{Bio. Sc.}]_i$ distinguish scientists in the physical or biological sciences, and women in the social sciences are the omitted comparison group. Cohort fixed effects θ_t control for variation in the propensity to patent across birth cohorts. Since equation (7) includes two endogenous variables, $\mathbb{1}[\text{Phys. Sc.}]_i$ and $\mathbb{1}[\text{Bio. Sc.}]_i$, we estimate two first-stage equations with two field-choice instruments:

$$\mathbb{1}[\text{Phys. Sc.}]_i = \gamma_1 \times Z_{i,t}^{Phys.Sc.} + \gamma_2 \times Z_{i,t}^{Bio.Sc.} + \vartheta_t + \varepsilon_i \tag{8}$$

$$\mathbb{1}[\text{Bio. Sc.}]_i = \delta_1 \times Z_{i,t}^{Phys.Sc.} + \delta_2 \times Z_{i,t}^{Bio.Sc.} + \vartheta_t + \varepsilon_i \tag{9}$$

The research field that woman i enters is predicted by the leave-one-out average field choices $Z_{i,t}$ of all women who enter science in the same year as woman i. Both first-stage equations control for entry-year fixed effects ϑ_t; all remaining variables are defined as above in equation (7). Women in the social sciences form the omitted category. We calculate the leave-one-out average field choices as:

$$Z_{i,t}^{Phys.Sc.} = \frac{1}{N_t - 1} \sum_{j \neq i, t=t}^{J} \mathbb{1}[\text{Phys. Sc.}]_j \quad Z_{i,t}^{Bio.Sc.} = \frac{1}{N_t - 1} \sum_{j \neq i, t=t}^{J} \mathbb{1}[\text{Bio. Sc.}]_j \tag{10}$$

where N_t is the total number of women entering science in year t, and for each individual woman i, $Z_{i,t}$ measures the average share of all other women J entering the physical or biological sciences in year t, respectively.

Jointly estimating equations (7), (8), and (9) with 2SLS identifies the causal effect of entering the physical or biological sciences on the probability that woman i will become an inventor under the exclusion restriction that factors impacting the field choice of *other* female scientists do not determine whether woman i would have patented in a different field. Overidentifying restriction tests support this exclusion restriction: Intuitively, the leave-one-out instruments represent pull-factors that are drawing all women entering science in a given year into specific fields. This allows us to leverage WWII enlistments at the

field-level as an additional push-factor that moves women into high-enlistment fields.

Complementary specifications include additional sets of instruments: interaction terms between leave-one-out shares $Z_{i,t}$ and the field-level enlistment rate, so that we now have two instruments per field. We adjust the first-stage equations accordingly, with ψ_k denoting the share of enlisted scientists in field k:

$$\mathbb{1}[\text{Phys. Sc.}]_i = \gamma_1 Z_{i,t}^{Phys.Sc.} + \gamma_2 Z_{i,t}^{Bio.Sc.} + \gamma_3 Z_{i,t}^{Phys.Sc.} \times \psi_k + \gamma_4 Z_{i,t}^{Bio.Sc.} \times \psi_k + \vartheta_t + \varepsilon_i \quad (11)$$

$$\mathbb{1}[\text{Bio. Sc.}]_i = \delta_1 Z_{i,t}^{Phys.Sc.} + \delta_2 Z_{i,t}^{Bio.Sc.} + \delta_3 Z_{i,t}^{Phys.Sc.} \times \psi_k + \delta_4 Z_{i,t}^{Bio.Sc.} \times \psi_k + \vartheta_t + \varepsilon_i \quad (12)$$

With four instruments and two endogenous variables these specifications are overidentified. The exclusion restriction requires that the instruments affect the outcome variables only through the endogenous field variables. We test this assumption, using the additional instruments, by regressing the residuals from a regression of the outcome variable $\mathbb{1}[Inventor]_i$ on field choice against the additional instrument. Results from this overidentifying restrictions test confirm that at least some of the instruments are exogenous (Table 7, columns 5), with a Hansen J test statistic of 2.208 and a p-value of 0.332, and therefore correctly excluded from equation (7). Scientist-level IV estimates are robust to including controls for individual characteristics, including controls for PhD recipients and scientists working in industry (Tables A.2 and A.3).

Scientist-Level IV Estimates

Scientist-level 2SLS estimates suggest that entering the physical sciences rather than the social sciences increases the probability that a woman will become an inventor by 11.6% (Table 6, column 5). By comparison, choosing the biological sciences (which have much lower patenting rates than the physical sciences) rather than social sciences increases a women's probability of becoming an inventor only by 3.1% (Table 6, column 5).

IV estimates are only slightly larger than OLS estimates (at 11.4 and 3.0% for the physical sciences, respectively) suggesting that endogeneity in field choice is limited at the individual level. Leave-one-out share instrumental variables are highly predictive and produce a very strong first stage, with a Kleibergen-Paap F-statistic above 1,100.

7. COUNTERFACTUALS

Given current rates of change, Bell et al. (2019) (Figure III) predict that it will take 118 years to reach gender parity in invention.[10] In this section, we ask how much more quickly the innovation gender gap would close if more women entered patent-intensive research fields

[10] 18.2% of US inventors born in 1980 are women, relative to 6.8% of inventors born in 1940, implying a gain of 0.27 percentage points a year.

in the physical and biological sciences.

Counterfactual share of women among inventors in MoS (1956)

Using inventor biographies in the MoS (1956), we calculate the share of women among inventors born between 1921 and 1929. Strikingly, these data anticipate the linear trend that Bell et al. (2019) illustrate for the 1940-80 birth cohorts: 0.6% of inventors born in 1921 are women, compared to 2.3% percent in 1928 and 6.8% percent in 1940 (Figure 1).[11]

To estimate the counterfactual share of women among inventors born 1921-29, we apply a simple non-parametric reweighting scheme (DiNardo et al. 1996), using the share of inventors among female scientists:[12]

$$\overline{\text{Inventor}}_g = f(k,g) \times Pr(\text{Inventor}|k,g) \tag{13}$$

where $\overline{\text{Inventor}}_g$ denotes the share of scientists of gender g who have at least one patent, $f(k,g)$ is distribution of individuals of gender g across research fields k, and $Pr(\text{Inventor}|k,g)$ is the probability that an individual of gender g has at least one patent conditional on working in field k. This notation allows us to estimate counterfactual shares by – independently and jointly – varying the distribution of women across fields and the probability to invent conditional on entering a field. Equation (13) allows us to estimate the counterfactual share of female inventors if women entered patent-intensive research fields at the same rate as men. Specifically, we weigh the conditional inventor shares of women in the respective fields by the density $f'(k,g)$ which represents the distribution of men across k. We estimate counterfactual female inventor shares for 1) the distribution of men across the physical, biological, and social sciences (Figure 12), and 2) the distribution of men across research fields determined by our k-means clustering algorithm (Figure 13).

First, we estimate the effect of changing women's distribution across fields while holding the probability that a woman has at least one patent constant within fields. Thus, we report counterfactual inventor shares among women by changing $f(k,g)$ while holding $Pr(\text{Inventor}|k,g)$ constant at observed levels. If women were distributed across the physical, biological, and social sciences in the same way as men, and the share of women who patent was constant within sciences, the share of female scientists who are inventors would be 5.5%, an increase of 51.5% over the observed inventor share of 3.6%. If women were distributed across (k-means) fields in the same way as men, the share of women who are inventors would be 8.3%, an increase of 127.8%.

[11] The average increase in the share of women among inventors, measured as OLS slope coefficients, is 0.23 percentage points per year for the MoS data, 0.27 percentage points per year for Bell et al. (2019), and 0.31 for the joint series combining both sources.

[12] We use the simplest reweighting scheme in DiNardo et al. (1996). They produce a counterfactual wage density for the case that unionization levels had not changed between 1979 and 1988. Their counterfactual density "is obtained by replacing the relative weight of the densities of unionized vs. nonunionized workers of 1988 by that of 1979" (p.11).

Next, we estimate counterfactuals allowing new female entrants to patent at a *different* rate from women who already are in a field. These counterfactuals use IV estimates of the effects of women joining fields on their probability to patent. Field-level estimates imply that 19.5% of women entering the physical sciences and 4.9% of women entering the biological sciences become inventors. Using these estimates, we assign 19.5% of women moving from the social to the physical sciences as additional inventors, 4.9% of women moving from the social to the biological sciences, and 14.6% of women moving from the biological to the physical sciences.[13]

Using field-level IV estimates to predict rates of invention for women entering the physical and biological sciences, the share of female inventors would be 105.8% higher if women were distributed across the physical, biological, and social sciences in the same way as men. Using IV estimates from the individual-level analysis, where we assign 11.6% of women moving from the social to the physical sciences, 3.0% of women moving from the social to the biological sciences, and 8.6% of women moving from the biological to the physical sciences as additional inventors, the counterfactual share of female inventors would be 62.4% higher.

Counterfactual share of women among inventors born between 1940 and 1980

To calculate the counterfactual innovation gender gap, we combine the counterfactual estimates of the share of women who are inventors with the augmented timeline of gender parity in Figure 1. Specifically, we implement the counterfactual inventor shares by assuming that the share of women among inventors in the 1921 to 1929 birth cohorts would have been at the respective counterfactual female inventor share. We then calculate the counterfactual innovation gender gap and the time to gender parity over the 1940-80 birth cohorts by extending a growth path from the counterfactual 1929 level: What would the female inventor share have been in each birth cohort between 1930 and 1980 if the share of women among inventors had grown at the growth rate implied by counterfactual levels for the 1921-29 birth cohorts?

We calculate counterfactual shares of female inventors for the 1930-80 birth cohorts using all four counterfactual scenarios: (1) *equity in the allocation of talent across sciences*, reweighting female scientists across sciences and keeping conditional patent rates constant, (2) *equity in the allocation of talent across fields*, reweighting female scientists across k-means fields and keeping conditional patent rates constant, (3) *equity in the allocation of talent across sciences and in patenting across fields*, reweighting female scientists across sciences and adjusting conditional patent rates by field-level IV estimates, and (4) *equity in the allocation of talent across sciences and in patenting at the individual level*, reweighting female scientists across sciences and adjusting patent rates by individual-level IV estimates.

[13] Because we are reweighing to the distribution of men, the counterfactual density has fewer women in the social sciences and all moves are out of the social and biological sciences.

If women were as likely to work in the physical and biological sciences as men, the share of female inventors born in 1980 would have been 15.9% higher (counterfactual share of 21.1% relative to the observed share of 18.2%, Figure 14, counterfactual 1, equity in the allocation of talent across sciences).

If women were distributed across (k-means) research fields as men, the share of women among inventors would be 70.9% higher (counterfactual share of 31.1%, Figure 15, counterfactual 2, equity in the allocation of talent across fields).

If we use field-level IV estimates for patenting, the share of women among inventors born in 1980 would be 55.2% higher (counterfactual share of 28.2%, Figure 16, counterfactual 3, equity in the allocation of talent across sciences and in patenting across fields).

If we use individual-level IV estimates for patenting, the share of women among inventors born in 1980 would be 23.9% higher (counterfactual share of 22.5%, Figure 17, counterfactual 4, equity in the allocation of talent across sciences and in patenting at the individual level).

Counterfactual time to gender parity: first birth cohort to reach 50% women among inventors

Using the counterfactual share of women inventors in 1980, we can calculate the remaining time to gender parity. Counterfactual estimates imply that changes towards a more equitable allocation of talent could help to close the innovation gender gap up to 81 years earlier than current estimates imply (Figure 18). If women were as likely to work in the physical and biological sciences as men (equity in the allocation of across sciences), the first birth cohort to reach parity would be 2064 (34 years earlier than implied by current estimates in Bell et al. (2019)). If women were distributed across fields like men (counterfactual 2, equity in the allocation of talent across fields), the first cohort to reach gender parity would be 2017 (81 years earlier). With equity in the allocation of talent across science and in patenting across fields (counterfactual 3, using field-level IV estimates to predict patenting), the first cohort to reach gender parity would be 2027 (73 years earlier). With equity in the allocation of talent across sciences and in patenting at the individual level (counterfactual 4, using individual-level IV estimates to predict patenting), the first cohort to reach gender parity would be 2055 (43 years earlier).

How many additional female inventors?

Counterfactuals predict the entry of additional women into innovation. The total increase in the number of women inventors, however, also depends on the effects of female entry on the total population of inventors: Do female inventors displace male inventors or does female entry enlarge the overall inventor population? To quantify how many more women enter depending on the replacement rates of male inventors, we decompose counterfactual entry:

$$r_{t,f'} \times (N_t + \Delta_t) = r_{t,f} N_t + s(r_{t,f'} - r_{t,f}) N_t + (1-s) \frac{(r_{t,f'} - r_{t,f})}{1 - r_{t,f'}} N_t \tag{14}$$

Here, $r_{t,f'}$ is the counterfactual share of female inventors in birth cohort t. N_t is the observed population in birth cohort t and Δ_t is its counterfactual increase. The substitution rate between male and female inventors s captures the share of female entrants that displace male inventors. The right-hand side of equation (14) adds three terms: $r_{t,f} N_t$, the observed count of female inventors per birth cohort, $(r_{t,f'} - r_{t,f}) N_t$ the number of female entrants who replace male inventors, as a share of the total inventor population N_t, and $\frac{(r_{t,f'} - r_{t,f})}{1 - r_{t,f'}} N_t$, the number of female entrants whose entry into innovation does not replace male inventors and therefore expands the population of inventors. Equation (14) introduces the substitution rate s which linearly combines replacement and female entry without replacement.

The total increase in the number of inventors in the economy depends on the *pure gain* in female inventors, the number of women who enter without displacing men. We define Δ_t as the counterfactual increase if $s = 0$, when female entrants do not replace men. Then, the counterfactual share of women becomes

$$r_{t,f'} = \frac{r_{t,f} N_t + \Delta_t}{N_t + \Delta_t} \tag{15}$$

Solving for Δ_t yields

$$\Delta_t = \frac{(r_{t,f'} - r_{t,f}) N_t}{1 - r_{t,f'}} \tag{16}$$

$(N_t + \Delta_t)$ is the total counterfactual population of inventors. $r_{t,f}$ is the observed share of female inventors per cohort and N_t is the number of inventors born in the United States each year between 1940 and 1980 from Bell et al. (2019).[14]

Figure 19 presents estimates for a range of values of s in counterfactual (3), equity in distribution and patenting across fields. The number of female inventors would increase by 86,057 (60.5%) if substitution is low (s = 0.01, i.e. one out of 100 female entrants replaces a man) and 67,814 (47.7%), if substitution is high (s = 1, i.e. every female entrant replaces a man).

How would this shift change aggregate invention?

Counterfactuals indicate that female entry into STEM would substantially increase the number of female inventors. How many inventions would these women generate? To answer this question, we decompose the net gain in patents by inventors of birth cohort t under the counterfactual density f':

[14] In total, 1.17 million inventors were born in the US between 1940 and 1980.

$$\text{Add. Patents}_{t,f'} = (r_{t,f'} - r_{t,f})N_t \left(\frac{1 - sr_{t,f'}}{1 - r_{t,f'}} \right) \times \text{Patents}_{t,w} - s(r_{t,f'} - r_{t,f})N_t \times \text{Patents}_{t,m} \quad (17)$$

On the right-hand side, the first term, $(r_{t,f'} - r_{t,f})N_t \left(\frac{1-sr_{t,f'}}{1-r_{t,f'}} \right) \times \text{Patents}_{t,w}$, captures patents by female entrants at the substitution rate s. The second term, $s(r_{t,f'} - r_{t,f})N_t \times \text{Patents}_{t,m}$, captures the lost patents by men who are replaced by female inventors. $\text{Patents}_{t,w}$ and $\text{Patents}_{t,m}$ capture the observed mean of patents granted to women and men inventors in birth cohort t between 1940-80.[15] Intuitively, the counterfactual change in patents reflects the difference in patents per person by male and female inventors (at observed levels). Since male inventors patent more, the net change in patenting will be negative if the substitution rate s is sufficiently high.[16]

Figure 20 reports the relative increase in patents resulting from female entry (counterfactual 3, equity in the allocation of talent across sciences and in patenting across fields) for different values for s. For inventors born in 1980, counterfactual patenting increases by 2.6% if substitution is low (s = 0.01), 0.2% more patents if substitution is moderate (s = 0.2) and reduces patenting if the substitution rate is higher than 0.21.

8. Roy (1951)-Model of Science

This section presents a simple Roy (1951) model to investigate distortions in three stages: entry into science, field choice, and patenting (Figure 21). In the first stage, individuals i choose to work in science or another job; the utility of working in another job is normalized to zero. In the second stage, scientists choose between research fields in the physical and social sciences.[17] In the third stage, individuals who specialized in STEM (the physical sciences) can choose to patent, while social scientists (e.g., in psychology or anthropology) do not patent.[18] Women account for half of the population ($\omega = 0.5$).

Talent, Wages, and Costs

Let $\phi_{f,i}$ denote individual i's talent. The index f takes three values, S, P, or N, where S denotes an individual's talent for the social sciences (and not patenting), while P denotes an individual's talent for patenting in the physical sciences, and N denotes the individual's talent for non-patent research in the physical sciences. Each individual is endowed with a vector

[15] Since patents per birth cohort for male and women are only available for patent issues between 1996 and 2012, data from Bell et al. (2019) suffer from censoring, and we report changes in patent as percentages rather than absolute numbers.

[16] Appendix Figure B7 plots the 'breakeven' substitution rate at which female entry results in a net gain in patenting. As the gender patent gap narrows, the breakeven substitution rate increases.

[17] The model abstracts from the biological sciences for ease of exposition.

[18] Data from the MoS (1956) show that patenting is rare in the social sciences: just 3.9% of scientists in a social science field ever file a patent, while 36.8% of scientists in the physical sciences are inventors.

$\Phi_i = \{\phi_{S,i}, \phi_{P,i}\, \phi_{N,i}\}$ of i.i.d. standard uniform talents across S, P, and N.

The distribution of talent for f is identical for women and men. An individual's talent in f determines both the wage they earn from specialising in f, and the cost of specialization c. After individuals see their talent draws, they choose whether to become a scientist, which field to enter, and whether to patent by maximizing their utility:

$$\max_f u_i = \max[w(\phi_{f,i}) - c(\phi_{f,i}), 0] \tag{18}$$

where $w(\phi_{f,i})$ is the wage function and c captures the cost of individual i working in f. The cost of specializing in f, $c(\phi_{f,i})$, for individual i is convex in their talent, so that more talented scientists face lower costs of entry: $c(\phi_{f,i}) = (1-\phi_{f,i})^2$. Wages increase additively in individual i's own talent in f and in the average talent of all people in f, but decrease as the field becomes more crowded and a larger share of individuals are active in f.

$$w(\phi_{f,i}) = \phi_{f,i} + \theta_f \mu_w^f + (1-\theta_f)\mu_m^f - \omega(1-F_w(\bar{f_w})) - (1-\omega)(1-F_m(\bar{f_m})) \tag{19}$$

μ_w^f and μ_m^f capture the average talent of female and male researchers in f. θ_f is the share of women among scientists in f.[19] $(1-F_w(\bar{f_w}))$ and $(1-F_m(\bar{f_m}))$ denote the share of women and men, respectively, who become scientists in f, given by the multivariate cumulative density function of Φ_i at $\bar{f}$, the minimum talent required to enter f. Intuitively, the share of entrants is given by the population share of individuals who have drawn their highest talent in f and whose talent in f exceeds the talent threshold in f. These individuals have a comparative advantage in f; it is determined jointly by the share of women who are above the talent threshold for women and by the share of men who are above the talent threshold for men $\omega(1-F_w(\bar{f_w})) + (1-\omega)(1-F_m(\bar{f_m}))$. While the population share of women ω is constant, the share of women in f, θ_f, varies with entry decisions. Appendix C derives the wage function under i.i.d. standard uniform talents.

[19] $\theta_f = \frac{(1-F_w(\bar{f_w}))}{(1-F_w(\bar{f_w}))+(1-F_m(\bar{f_m}))}$

Indifference conditions and model solution

We use the wage equations to define a system of simultaneous equations:

$$w(\bar{P}_w) - c(\bar{P}_w) = w(\bar{N}_w) - c(\bar{N}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{N}_m) - c(\bar{N}_m) \tag{20}$$

$$w(\bar{P}_w) - c(\bar{P}_w) = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{21}$$

$$w(\bar{N}_w) - c(\bar{N}_w) = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{N}_m) - c(\bar{N}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{22}$$

$$w(\bar{P}_w) - c(\bar{P}_w) = 0 \qquad w(\bar{P}_m) - c(\bar{P}_m) = 0 \tag{23}$$

$$w(\bar{N}_w) - c(\bar{N}_w) = 0 \qquad w(\bar{N}_m) - c(\bar{N}_m) = 0 \tag{24}$$

$$w(\bar{S}_w) - c(\bar{S}_w) = 0 \qquad w(\bar{S}_m) - c(\bar{S}_m) = 0 \tag{25}$$

We derive these equations by applying backward induction to Figure 21. Equations (20) - (22) describe indifference conditions of choosing between fields; equations (23) - (25) describe indifference conditions of choosing between becoming a scientist in a given field and the outside option.

Solving this system allows us to identify talent thresholds – the level of talent at which the marginal woman or man is indifferent about entering: $\bar{S}_w$, $\bar{S}_m$, $\bar{P}_w$, $\bar{P}_m$, $\bar{N}_w$, and $\bar{N}_m$. In equilibrium all individuals with comparative advantage and talent above the threshold enter and no individual with talent below the threshold enters (Appendix C).

Solving the model numerically yields the following thresholds.

$$\{\bar{P}_w, \bar{P}_m\} = \{0.20585, 0.20585\}$$
$$\{\bar{N}_w, \bar{N}_m\} = \{0.20585, 0.20585\}$$
$$\{\bar{S}_w, \bar{S}_m\} = \{0.20585, 0.20585\}$$

Talent thresholds are symmetrical across all fields and across genders (Figure 22). If women and men are equally talented and there are no distortions, men and women should be equally likely to enter science, equally likely to choose STEM, and equally likely to patent. This symmetrical solution is the unique subgame-perfect Nash equilibrium of our model (Appendix C).

Taking the Model to the Data: Misallocation in Science

Our model predicts that, in equilibrium, men and women are equally likely to become scientists, equally likely to be active in physical rather than social sciences, and equally likely to patent. Observed gender shares, however, contradict these predictions: 94% of active scientists are male; 74.3% of men are in physical sciences, compared with 50.2% of women, and 37.7% of men in the physical sciences patent, compared with 9.4% of women.

We use our model to quantify the distortions that women must face to generate these discrepancies. Specifically, we use men's participation in science, the physical sciences, and

patenting as a benchmark for the participation of women *without* distortions. This approach allows us to quantify distortions in entering science, working in the physical sciences, and in patenting.[20]

Entry Distortion

Only 6% of scientists in the MoS are female. How much higher must women's cost of becoming a scientist be to generate this disparity? To answer this question, we extend our model to include a cost parameter α that applies to all women's payoffs of being a scientist.

This parameter α captures bias that women face entering science in *any* field. For example, α may be high if women face higher barriers in earning graduate degrees (Goldin, Katz, and Kuziemko 2006),[21] or if women have reduced incentives to invest in their education because they expect to work for fewer years (Adda, Dustmann, and Stevens 2017) or because they anticipate having children (Kim and Moser 2024).[22] To quantify entry distortions, we add the observed gender share of researchers as an equilibrium condition and let the model set $\bar{P}_w, \bar{P}_m, \bar{N}_w, \bar{N}_m, \bar{S}_w, \bar{S}_m$, and α to match the data.

We define the following modified system of simultaneous equations:

$$
\begin{aligned}
w(\bar{P}_w) - c(\bar{P}_w) - \alpha &= w(\bar{N}_w) - c(\bar{N}_w) - \alpha & \qquad w(\bar{P}_m) - c(\bar{P}_m) &= w(\bar{N}_m) - c(\bar{N}_m) \\
w(\bar{P}_w) - c(\bar{P}_w) - \alpha &= w(\bar{S}_w) - c(\bar{S}_w) - \alpha & \qquad w(\bar{P}_m) - c(\bar{P}_m) &= w(\bar{S}_m) - c(\bar{S}_m) \\
w(\bar{N}_w) - c(\bar{N}_w) - \alpha &= w(\bar{S}_w) - c(\bar{S}_w) - \alpha & \qquad w(\bar{N}_m) - c(\bar{N}_m) &= w(\bar{S}_m) - c(\bar{S}_m) \\
w(\bar{P}_w) - c(\bar{P}_w) - \alpha &= 0 & \qquad w(\bar{P}_m) - c(\bar{P}_m) &= 0 \\
w(\bar{N}_w) - c(\bar{N}_w) - \alpha &= 0 & \qquad w(\bar{N}_m) - c(\bar{N}_m) &= 0 \\
w(\bar{S}_w) - c(\bar{S}_w) - \alpha &= 0 & \qquad w(\bar{S}_m) - c(\bar{S}_m) &= 0
\end{aligned}
$$

$$\text{Share of Women Scientists} = 0.06$$

Numerically solving the model with entry distortions yields:

[20] For simplicity, we assume that all information about misallocation that individuals consider in their occupational choice is contained in their expectation of talent threshold values. We assume that individuals have perfect information with regards to the thresholds.

[21] Goldin et al. (2006) show that – conditional on high school-class rank, IQ, high school-class choice, and family background – women were 7.5 percentage points less likely than comparable men to graduate four-year college in 1957 and 2.5 percentage points less likely in 1972.

[22] Since the specific value of α depends on our choice of the wage function, we focus *relative* differences in minimum talent thresholds for men and women

$$\{\bar{P}_w, \bar{P}_m\} = \{0.97832, 0.14376\}$$
$$\{\bar{N}_w, \bar{N}_m\} = \{0.97832, 0.14376\}$$
$$\{\bar{S}_w, \bar{S}_m\} = \{0.97832, 0.14376\}$$
$$\{\alpha\} = \{1.56723\}$$

A positive value of α indicates that women face higher costs of becoming scientists than men. Talent thresholds imply that women face a 5.8 times higher talent threshold to enter science than men.[23] This result is consistent with a "Marie Curie strategy" (Rossiter 1982, p.159):

> "to be considered 'equal' to men, (...) women had to be 'better'. The most determined of them (the 'survivors') therefore internalized this double standard and settled down to becoming 'Madame Curies': earning an extra degree was just the first step in a lifelong struggle to overcompensate for being women in a man's occupation."

If women face higher thresholds entering science, they should be positively selected. Data on gender differences in human capital acquisition confirm this prediction.[24] Female scientists are significantly more likely to hold a PhD than men: 64% of women have a doctorate, compared with 58% of men (p-value = 0.00). Moreover, they are significantly more likely to have graduated from elite Ivy+ universities: 32% of women have an Ivy+ degree, compared with 28% of men (p-value = 0.00).[25]

Confirming the predictions of our model, women who entered during WWII are less positively selected. For every ten additional female entrants, the average share of women with a PhD in an entry cohort is 1 percentage point lower (Figure 11). The average share of women with an Ivy+ degree is 0.3 percentage points lower per ten additional entrants (Figure B6).

Figure 23 illustrates another important prediction of our model: Wages increase with the size of entry distortions, encouraging less talented men to enter science. Higher talent thresholds for women discourage female entry and reduce competition for jobs. Moreover, the entry of highly selected female scientists raises the average level of talent. Both forces raise wages in science, and encourage men who would be below the talent threshold without distortions to enter science.

[23] Given by the relative size of male and female entry thresholds: $\frac{0.97832}{0.14376} - 1 = 5.8$.

[24] We focus on education as a measure of selection rather than research output to account for distortions women face in output markets, such as patenting and publishing.

[25] These differences hold for both scientists in academia and in industry employment. 65% of women in academia have a PhD and 34% have an Ivy+ degree, while 59% of men have a PhD and 27% have an Ivy+ degree. In industry employment, 59% of women have a PhD (52% of men) and 20% of women hold an Ivy+ degree (14% of men).

Field Distortion

In addition to entry into science, data on the fields of female and male scientists indicate significant gender differences in the allocation of scientific talent across fields: Women are 32% less likely to be in the physical sciences then men.[26] What are the distortions that drive this additional disparity? To answer this question, we introduce an additional cost parameter β that women face when they do research in the physical sciences. For instance, β may be high if face stereotypes about their aptitude for specific research tasks (see Kahn and Ginther (2017) for a review) or distortions that relate to professional norms (Bertrand, Goldin, and Katz 2010; Goldin 2014; Cortes and Pan 2016; Goldin and Katz 2016), scheduling flexibility (Bronson 2015), teacher discrimination, role models (Carrell, Page, and West 2010) and work environment concerns (Wiswall and Zafar 2018).[27] Including the relative field distribution of male and female scientists as an additional equilibrium condition, we let the model set talent thresholds, α and β to match the data (Appendix C).

Specifically, we derive the following modified system of simultaneous equations:

$$
\begin{aligned}
w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta &= w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta & \quad w(\bar{P}_m) - c(\bar{P}_m) &= w(\bar{N}_m) - c(\bar{N}_m) \\
w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta &= w(\bar{S}_w) - c(\bar{S}_w) - \alpha & \quad w(\bar{P}_m) - c(\bar{P}_m) &= w(\bar{S}_m) - c(\bar{S}_m) \\
w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta &= w(\bar{S}_w) - c(\bar{S}_w) - \alpha & \quad w(\bar{N}_m) - c(\bar{N}_m) &= w(\bar{S}_m) - c(\bar{S}_m) \\
w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta &= 0 & \quad w(\bar{P}_m) - c(\bar{P}_m) &= 0 \\
w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta &= 0 & \quad w(\bar{N}_m) - c(\bar{N}_m) &= 0 \\
w(\bar{S}_w) - c(\bar{S}_w) - \alpha &= 0 & \quad w(\bar{S}_m) - c(\bar{S}_m) &= 0
\end{aligned}
$$

$$\text{Share of Women Scientists} = 0.06$$

$$\frac{\text{Women in Physical Sciences}}{\text{Men in Physical Sciences}} = 0.68$$

Numerically solving the model with both entry and field distortions yields the following talent thresholds and cost parameters:

[26] See Figure 10 for the full distribution of men and women across physical, biological, and social sciences. For simplicity, we only consider the physical and social sciences in our model. Of all women who are in the physical or social sciences, 50.2% are in the physical sciences – relative to 74.3% of men. We calculate the relative probability of women entering the physical sciences accordingly as $\frac{0.502}{0.743} = 0.676$.

[27] See Patnaik, Wiswall, and Zafar (2021) for a review of wage and non-wage determinants of gender gaps in college major choice.

$$\{\bar{P}_w, \bar{P}_m\} = \{0.98558, 0.14408\}$$
$$\{\bar{N}_w, \bar{N}_m\} = \{0.98558, 0.14408\}$$
$$\{\bar{S}_w, \bar{S}_m\} = \{0.96347, 0.14347\}$$
$$\{\alpha, \beta\} = \{1.55229, 0.02161\}$$

While the absolute differences in the thresholds are small, the relative impact of field distortions is large: The model with entry distortions suggest that only the most talented 2.17 percent of women enter a given field. Field distortions further restrict entry into the physical sciences to the top 1.44 percent of women, a reduction of 34%.

Patenting Distortion

Even conditional on entering science and working in the same fields as men, women patent less. Just 9.4% of women in the physical sciences patent, compared with 37.7% of men. To quantify the distortions that generate these gender differences in patenting (conditional on entry into science and working in a certain field) *within fields*, we introduce a third cost parameter γ that affects women's payoffs from patenting - conditional on entering the physical sciences.

Intuitively, γ may capture bias relating to norms, work environments, and similar barriers that may keep women from patenting. Moreover, γ may also capture discrimination and disparities in the patent system (Pairolero, Toole, DeGrazia, Teodorescu, and Pappas 2022), e.g. if patent examiners value the work of women less or if institutions (like universities or firms) do not let women partake in the gains from patenting. Introducing the relative patenting rate of male and female scientists within the physical sciences as an additional equilibrium condition, we let the model set talent thresholds, α, β, and γ to match the data (Appendix C).

Specifically, we derive the following modified system of simultaneous equations:

$$\begin{aligned}
w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta - \gamma &= w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta & w(\bar{P}_m) - c(\bar{P}_m) &= w(\bar{N}_m) - c(\bar{N}_m) \\
w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta - \gamma &= w(\bar{S}_w) - c(\bar{S}_w) - \alpha & w(\bar{P}_m) - c(\bar{P}_m) &= w(\bar{S}_m) - c(\bar{S}_m) \\
w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta &= w(\bar{S}_w) - c(\bar{S}_w) - \alpha & w(\bar{N}_m) - c(\bar{N}_m) &= w(\bar{S}_m) - c(\bar{S}_m) \\
w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta - \gamma &= 0 & w(\bar{P}_m) - c(\bar{P}_m) &= 0 \\
w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta &= 0 & w(\bar{N}_m) - c(\bar{N}_m) &= 0 \\
w(\bar{S}_w) - c(\bar{S}_w) - \alpha &= 0 & w(\bar{S}_m) - c(\bar{S}_m) &= 0
\end{aligned}$$

Share of Women Scientists $= 0.06$

$$\frac{\text{Women in Physical Sciences}}{\text{Men in Physical Sciences}} = \frac{0.502}{0.743}$$
$$\frac{\text{Women Inventors}}{\text{Men Inventors}} = \frac{0.094}{0.377}$$

Numerically solving the model with entry, field, and patenting distortions results in the following minimum talent thresholds and cost parameters:

$$\{\bar{P}_w, \bar{P}_m\} = \{0.99643, 0.14478\}$$
$$\{\bar{N}_w, \bar{N}_m\} = \{0.97448, 0.14365\}$$
$$\{\bar{S}_w, \bar{S}_m\} = \{0.96347, 0.14347\}$$
$$\{\alpha, \beta, \gamma\} = \{1.55229, 0.01123, 0.01952\}$$

These results indicate that only the 0.36% most talented women patent in the physical sciences.

How much would reducing distortions increase the number of female inventors?

Next, we estimate the counterfactual increase in the number of female inventors that would result from removing distortions (Table 8). While entry distortions explain the largest absolute differences in female participation in science, barriers between fields have important *relative* effects. Specifically, we calculate the effect of removing a given distortion on the number of female inventors by comparing the share of individuals who would be talented enough to enter an area of science at each resulting minimum talent threshold.[28] Importantly, the effects of separate distortions compound throughout the three stages of our model. Consequently, reducing multiple distortions has a disproportionally larger effect on the number women inventors than addressing distortions one-by-one.

First, we consider three counterfactual scenarios: *equity in the allocation of talent across sciences* would increase the number of female inventors by 19.40%, *equity in the allocation of talent across sciences and in patenting across fields* would increase the number of female inventors by 66.38%, and *equity in the allocation of talent across sciences and in patenting at the individual level* would increase the female inventor count by 29.65%. Jointly accounting for changes in women's field choices and patenting rates yields the increase in the number of female inventors (Figure 24).

Second, we consider the effect of removing distortions entirely. The largest counterfactual increase in female inventors is associated with removing both field and

[28] As before, the share of entrants for threshold $\bar{F}$ in f is given by $\frac{(1-\bar{F}^3)}{3}$. We derive entrant shares under i.i.d. uniform talents in Appendix C.

patenting distortions. Without these distortions, the number of female inventors would increase by 497.05%. Only removing patenting distortions would increase the number of female inventors by 300.02%. Only removing field distortions would increase the number of female inventors by 19.40% (this scenario is equivalent to *equity in the allocation of talent across sciences*).

Finally, we use our model to study changes in female participation in science during WWII. To investigate the effects of WWII, we calibrate the model separately for each cohort of scientists starting their career between 1933 and 1951 and obtain year-by-year estimates of talent thresholds and distortions. These estimates allow us to disentangle the effects of WWII on women entering science and patenting.

We find that all three talent thresholds fall after 1942 (Figure B8), as more women entered science. Notably, the largest change in talent thresholds for entry into the physical sciences (0.943 of its 1933 level), consistent with reduced barriers to entry in these fields. WWII increases female innovation by reducing patenting distortions within fields: While patenting thresholds remain high (at 0.993), women entering the physical sciences during WWII are 1.97 times more likely to become inventors than women entering in 1933.[29]

9. Conclusion

Analyzing the biographies of American scientists – linked with their patents – we have found that differences in the allocation of talent across research fields are a major driver of the innovation gender gap: Women patent less because they are more likely to work in fields that are less intensive in innovation, producing fewer patents.

Using the WWII draft of male scientist as an instrument for female entry, we ask whether shifts towards a more equitable distribution of scientists across fields could help close the innovation gender gap. We find that bringing more women into patent-intensive STEM fields in the biological and physical sciences would lead to a significant increase in female innovation. Field-level IV estimates imply that, for every 5 additional women entering the physical sciences, one additional woman would become an inventor.

Incorporating these estimates in a series of counterfactuals, we find that – if women entered STEM at the same rate as men – 61% more American women born 1940-80 would have become inventors, and women would have produced roughly 3% more patents per birth cohort. With gender equity in the allocation of talent and in patenting, the patent gap would close in 37 years, 81 years earlier than is implied by current estimates.

We evaluate these effects through the lens of a three-stage Roy model that allows us to investigate distortions at the level of entry into science, a scientist's choice of research fields, and patenting. Taking the model to the data, we find that women have to be substantially more

[29] Thresholds imply that, in 1933, 0.235 percent of the population of women would have been talented enough to become scientists, relative to 0.697 percent in 1945.

talented than men to enter science, more talented to enter the physical sciences, and more talented to patent. While this paper has shown that improvements in the allocation of talent *within* science can help to close the innovation gender gap, our results also emphasize the need to better understand the barriers that prevent talented women from entering science in the first place.

References

ACEMOGLU, D., D. H. AUTOR, AND D. LYLE (2004): "Women, war, and wages: The effect of female labor supply on the wage structure at midcentury," *Journal of Political Economy*, 112, 497–551.

ADDA, J., C. DUSTMANN, AND K. STEVENS (2017): "The career costs of children," *Journal of Political Economy*, 125, 293–337.

BELL, A., R. CHETTY, X. JARAVEL, N. PETKOVA, AND J. VAN REENEN (2019): "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 134, 647–713.

BERTRAND, M., C. GOLDIN, AND L. F. KATZ (2010): "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American Economic Journal: Applied Economics*, 2, 228–255.

BIASI, B. AND P. MOSER (2021): "Effects of copyrights on science: Evidence from the WWII book republication program," *American Economic Journal: Microeconomics*, 13, 218–260.

BRONSON, M. A. (2015): "Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women," *Unpublished manuscript*, 2.

CARRELL, S. E., M. E. PAGE, AND J. E. WEST (2010): "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 125, 1101–1144.

CATTELL, J. (1956): *American Men of Science: A Biographical Directory*.

CATTELL, J. M. (1906): "A statistical study of American men of science: The selection of a group of one thousand scientific men," *Science*, 24, 658–665.

CHEN, A. AND P. MOSER (2024): "Shell Shock: Effects of Combat on American Scientists after WWII," *Working Paper*.

CORTES, P. AND J. PAN (2016): "Prevalence of Long Hours and Skilled Women's Occupational Choices," *IZA Discussion Paper*.

DINARDO, J., N. M. FORTIN, AND T. LEMIEUX (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

EUROPEAN COMMISSION (2021): *She Figures 2021 – Gender in Research and Innovation – Statistics and Indicators*, Publications Office.

FERNÁNDEZ, R., A. FOGLI, AND C. OLIVETTI (2004): "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 119, 1249–1299.

GOLDIN, C. (2014): "A grand gender convergence: Its last chapter," *American Economic Review*, 104, 1091–1119.

GOLDIN, C. AND L. F. KATZ (2016): "A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation," *Journal of Labor Economics*, 34, 705–746.

GOLDIN, C., L. F. KATZ, AND I. KUZIEMKO (2006): "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic Perspectives*, 20, 133–156.

GOLDIN, C. AND C. OLIVETTI (2013): "Shocking labor supply: A reassessment of the role of World War II on women's labor supply," *AEA Papers & Proceedings*, 103, 257–262.

GOLDIN, C. D. (1991): "The role of World War II in the rise of women's employment," *The American Economic Review*, 741–756.

GROSS, D. P. AND B. N. SAMPAT (2023): "America, jump-started: World War II R&D and the takeoff of the US innovation system," *American Economic Review*, 113, 3323–3356.

HSIEH, C.-T., E. HURST, C. I. JONES, AND P. J. KLENOW (2019): "The allocation of talent and US economic growth," *Econometrica*, 87, 1439–1474.

IARIA, A., C. SCHWARZ, AND F. WALDINGER (2022): "Gender gaps in academia: Global evidence over the twentieth century," *Available at SSRN 4150221*.

JAWORSKI, T. (2014): ""You're in the Army Now:" The Impact of World War II on Women's Education, Work, and Family," *The Journal of Economic History*, 74, 169–195.

KAHN, S. AND D. GINTHER (2017): "Women and STEM," *NBER Working Paper*.

KIM, S. D. AND P. MOSER (2024): "Women in Science. Lessons from the Baby Boom," *Working Paper*.

LEVENSHTEIN, V. I. ET AL. (1966): "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, Soviet Union, vol. 10, 707–710.

MINCER, J. (1978): "Family migration decisions," *Journal of Political Economy*, 86, 749–773.

MOSER, P. AND S. PARSA (2024): "McCarthy and the Red-ucators: Effects of Political Persecution on Science," *Working Paper*.

MOSER, P., S. PARSA, AND S. SAN (2024): "Immigration, Science, and Invention. Lessons from the Quota Acts," *Working Paper*.

OECD (2023): *Inventors (indicator)*, Accessed on 18 November 2023.

PAIROLERO, N., A. TOOLE, C. DEGRAZIA, M. H. TEODORESCU, AND P.-A. PAPPAS (2022): "Closing the gender gap in patenting: evidence from a randomized control trial at the USPTO," in *Academy of Management Proceedings*, Academy of Management Briarcliff Manor, NY 10510, vol. 2022, 14401.

PATNAIK, A., M. WISWALL, AND B. ZAFAR (2021): "College majors 1," *The Routledge handbook of the economics of education*, 415–457.

PHILIPS, L. (2000): "The double metaphone search algorithm," *C/C++ users journal*, 18, 38–43.

ROSE, E. K. (2018): "The rise and fall of female labor force participation during World War II in the United States," *The Journal of Economic History*, 78, 673–711.

ROSSITER, M. W. (1982): *Women Scientists in America: Struggles and Strategies to 1940*, vol. 1, JHU Press.

ROY, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

SINHA, A., Z. SHEN, Y. SONG, H. MA, D. EIDE, B.-J. HSU, AND K. WANG (2015): "An overview of microsoft academic service (mas) and applications," in *Proceedings of the 24th international conference on world wide web*, 243–246.

STOILOS, G., G. STAMOU, AND S. KOLLIAS (2005): "A string metric for ontology alignment," in *The Semantic Web–ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005. Proceedings 4*, Springer, 624–637.

WISWALL, M. AND B. ZAFAR (2018): "Preference for the workplace, investment in human capital, and gender," *The Quarterly Journal of Economics*, 133, 457–507.

Table 1: Summary Statistics for Female and Male Scientists

	Full Sample		Male Scientists		Female Scientists		Difference	
Variable	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Female - Male	
Inventor	0.221	0.415	0.232	0.422	0.036	0.187	-0.196	***
Publications	16.643	26.708	16.999	27.106	10.807	18.055	-6.191	***
Ivy+ Education	0.239	0.427	0.234	0.424	0.321	0.467	0.087	***
PhD	0.580	0.493	0.577	0.494	0.638	0.481	0.061	***
Academic Job	0.754	0.431	0.746	0.435	0.877	0.329	0.131	***
Industry Job	0.246	0.431	0.254	0.435	0.123	0.329	-0.131	***
Physical Sciences	0.505	0.500	0.518	0.500	0.292	0.455	-0.226	***
Biological Sciences	0.310	0.463	0.304	0.460	0.419	0.493	0.115	***
Social Sciences	0.185	0.388	0.179	0.383	0.290	0.454	0.111	***
Parent	0.711	0.454	0.740	0.438	0.221	0.415	-0.519	***
Number of Children	1.618	1.361	1.691	1.350	0.414	0.884	-1.277	***
Married	0.817	0.387	0.843	0.364	0.388	0.487	-0.455	***
Birth Year	1909	12.372	1909	12.351	1906	12.396	-2.865	***
Career Start Year	1934	12.744	1934	12.746	1932	12.509	-2.391	***
Field Male Inventor Rate	0.227	0.178	0.231	0.179	0.163	0.142	-0.068	***
N	70,039		66,019		4,020			

Note: Physical, Biological, and Social Sciences refer to the three MoS (1956) volumes. Scientists' fields are determined by a k-means clustering algorithm that leverages key words describing each individual scientist's research ($k = 100$). Field male inventor rate measures the leave-one-out share of male scientists in the same k-means field who are inventors for each scientist. Inventors are defined as scientists who file at least one patent during their career. Difference column reports results from two-sample t-tests.
*** p < 0.01 ** p < 0.05 * p < 0.1

Table 2: OLS Decomposition of the Gender Innovation Gap

	(1) Baseline	(2) Publications	(3) PhD	(4) Industry Job	(5) Married	(6) Parent	(7) Physical or Bio Sciences	(8) All Controls
Female	-0.194***	-0.192***	-0.166***	-0.155***	-0.150***	-0.155***	-0.043***	0.008
	(0.003)	(0.004)	(0.005)	(0.003)	(0.005)	(0.005)	(0.003)	(0.006)
Publications		0.000***						0.000***
		(0.000)						(0.000)
Female × Publications		-0.000						0.000
		(0.000)						(0.000)
PhD			0.049***					0.067***
			(0.003)					(0.003)
Female × PhD			-0.050***					-0.060***
			(0.007)					(0.007)
Industry				0.193***				0.121***
				(0.004)				(0.004)
Female × Industry				-0.132***				-0.060***
				(0.014)				(0.013)
Married					0.049***			-0.004
					(0.004)			(0.005)
Female × Married					-0.057***			0.004
					(0.007)			(0.009)
Children						0.048***		0.031***
						(0.004)		(0.004)
Female × Children						-0.063***		-0.043***
						(0.008)		(0.009)
Physical Science							0.334***	0.295***
							(0.003)	(0.003)
Female × Phys. Sc.							-0.237***	-0.199***
							(0.009)	(0.009)
Biological Science							0.057***	0.030***
							(0.003)	(0.003)
Female × Bio. Sc.							-0.035***	-0.008
							(0.005)	(0.005)
Constant	0.232***	0.226***	0.204***	0.184***	0.191***	0.196***	0.042***	-0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)
Birth year FE	✓	✓	✓	✓	✓	✓	✓	✓
Birth state FE	✓	✓	✓	✓	✓	✓	✓	✓
Mean Y	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221
Observations	70,033	70,033	70,033	70,033	70,033	70,033	70,033	70,033
R-squared	0.022	0.022	0.025	0.060	0.023	0.024	0.145	0.165

Note: Robust errors in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.1

Table 3: Field-Level TWFE Estimates of the Effects of Enlistment on Female Entry

	(1) Female Entrants	(2) Female Entrants	(3) Female Entrants	(4) Female Entrants	(5) IHS Female Entrants
Above Median Enlistment (1/0)	0.349**				
	(0.133)				
Enlistment Share (cont.)		3.861***		4.056***	1.750***
		(0.887)		(0.687)	(0.324)
2nd Quartile Enlistment			-0.113		
			(0.216)		
3rd Quartile Enlistment			0.119		
			(0.130)		
Top Quartile Enlistment			0.466***		
			(0.126)		
Model	OLS	OLS	OLS	PQML	OLS
Year FE	✓	✓	✓	✓	✓
Field FE	✓	✓	✓	✓	✓
Mean Y	0.924	0.924	0.924	0.983	0.541
Observations	3,000	3,000	3,000	2,820	3,000
R-squared	0.763	0.764	0.764	–	0.593

Note: OLS and Poisson estimates of two-way fixed effects difference-in-differences specifications. Treatment defined as $\mathbb{1}[Enlistment > Median]$ in Column (1) and as the share of enlisted male scientists in Columns (2), (4), and (5). Treatment defined along quartiles of the enlistment intensity distribution across *k*-means fields, with the bottom quartile serving as control group, in Column (3). Robust errors in parentheses. Poisson quasi-maximum likelihood specifications in Column (4) exclude 6 fields that never hire women throughout our sample. Column (5) reports estimates using an inverse hyperbolic sine-transformed outcome variable.
*** p < 0.01 ** p < 0.05 * p < 0.1

Table 4: Field-Level IV Estimates for Women Entering the Physical Sciences

Y =	(1) Women Inventors	(2) Women Inventors	(3) Women Scientists	(4) Women Inventors
	OLS	Reduced Form	First Stage	IV
Female Scientists	0.095***			0.195***
	(0.022)			(0.070)
Predicted Female Scientists		0.034**	0.175***	
		(0.013)	(0.032)	
Year FE	✓	✓	✓	✓
Field FE	✓	✓	✓	✓
Mean Y	0.069	0.069	0.656	0.069
Observations	1,000	1,000	1,000	1,000
Cragg-Donald F				16.763
Kleibergen-Paap F				29.179

Note: IV regressions estimated with 2SLS. Sample limited to k-means fields in the physical sciences. Robust errors clustered by field in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.1

Table 5: Field-Level IV Estimates for Women Entering the Biological Sciences

Y =	(1) Women Inventors	(2) Women Inventors	(3) Women Scientists	(4) Women Inventors
	OLS	Reduced Form	First Stage	IV
Female Scientists	0.039***			0.049***
	(0.008)			(0.012)
Predicted Female Scientists		0.026***	0.536***	
		(0.005)	(0.042)	
Year FE	✓	✓	✓	✓
Field FE	✓	✓	✓	✓
Mean Y	0.030	0.030	1.188	0.030
Observations	975	975	975	975
Cragg-Donald F				122.879
Kleibergen-Paap F				162.531

Note: IV regressions estimated with 2SLS. Sample limited to k-means fields in the biological sciences. Robust errors clustered by field in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.1

Table 6: Individual-Level IV Estimates

Y =	(1) Inventor	(2) Inventor	(3) Physical Sciences	(4) Biological Sciences	(5) Inventor
	OLS	Reduced Form	First Stage	First Stage	IV
Physical Sciences (1/0)	0.114***				0.116***
	(0.024)				(0.025)
Bio. Sciences (1/0)	0.030***				0.031***
	(0.006)				(0.005)
Phys. Pull		-9.914***	-86.120***	3.384**	
		(2.178)	(1.717)	(1.512)	
Bio. Pull		-2.411***	0.897	-81.937***	
		(0.427)	(0.628)	(2.045)	
Year FE	✓	✓	✓	✓	✓
Mean Y	0.045	0.045	0.288	0.394	0.045
Observations	1,951	1,951	1,951	1,951	1,951
Cragg-Donald F					8632.690
Kleibergen-Paap F					1110.048

Note: IV regressions estimated with 2SLS. Inventor is an indicator for whether a scientist has at least one patent. Physical and Biological Sciences are indicator variables for scientists in the respective sciences. Sample limited to female scientists. Robust errors clustered by field in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.10

Table 7: Individual-Level IV Estimates and Overidentifying Restrictions Test

Y =	(1) Inventor	(2) Inventor	(3) Physical Sciences	(4) Biological Sciences	(5) Inventor
	OLS	Reduced Form	First Stage	First Stage	IV
Physical Sciences (1/0)	0.114***				0.116***
	(0.024)				(0.025)
Bio. Sciences (1/0)	0.030***				0.030***
	(0.006)				(0.005)
Phys. Pull		-10.124***	-84.182***	4.771***	
		(2.391)	(2.045)	(1.546)	
Bio. Pull		-2.841***	1.721**	-81.780***	
		(0.600)	(0.842)	(2.001)	
Phys. Pull × Enlistment		-0.813	-3.622***	-4.127***	
		(0.642)	(1.117)	(1.267)	
Bio. Pull × Enlistment		0.801	1.819**	2.573***	
		(0.494)	(0.743)	(0.821)	
Year FE	✓	✓	✓	✓	✓
Mean Y	0.045	0.045	0.288	0.394	0.045
Observations	1,951	1,951	1,951	1,951	1,951
Cragg-Donald F					4704.670
Kleibergen-Paap F					613.145
Hansen J					2.208
Hansen J (p-value)					0.332

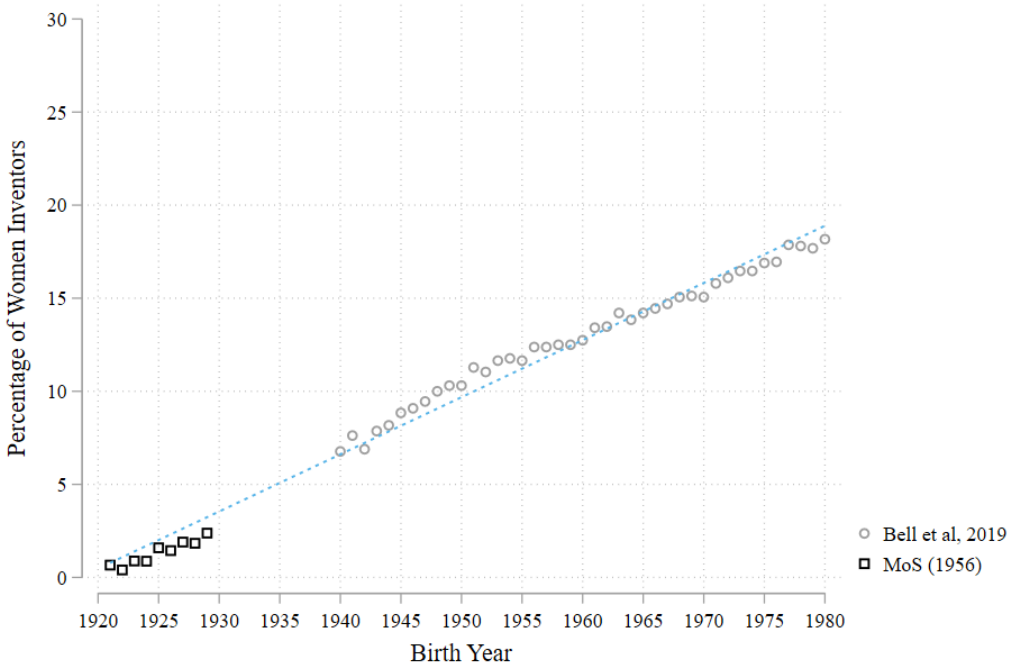
Note: IV regressions estimated with 2SLS. Inventor is an indicator for whether a scientist has at least one patent. Physical and Biological Sciences are indicator variables for scientists in the respective sciences. Enlistment is the share of enlisted draft-eligible scientists in the same k-means field. Robust errors clustered by field in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.10

Table 8: Additional Female Inventors from Removing Distortions

Counterfactual	$\bar{P}_m$	$\bar{P}_w$	$\bar{N}_m$	$\bar{N}_w$	$\bar{S}_m$	$\bar{S}_w$	Δ Female Inventors
Baseline	0.1442	0.9879	0.1436	0.9744	0.1436	0.9725	-
Counterfactual:							
Equity in the Allocation of Talent across Sciences	0.1447	0.9957	0.1435	0.9694	0.1435	0.9694	+19.40%
Equity in the Allocation of Talent across Sciences and in Patenting across Fields	0.1436	0.9703	0.1446	0.9941	0.1436	0.9703	+66.38%
Equity in the Allocation of Talent across Sciences and in Patenting at Individual Level	0.1447	0.9954	0.1435	0.9696	0.1435	0.9696	+29.65%
Fully Removing:							
Field Distortion	0.1447	0.9957	0.1435	0.9694	0.1435	0.9694	+19.40%
Patenting Distortion	0.1441	0.9856	0.1441	0.9856	0.1435	0.9635	+300.02%
Field Choice and Patenting Distortion	0.1438	0.9783	0.1438	0.9783	0.1438	0.9783	+497.05%

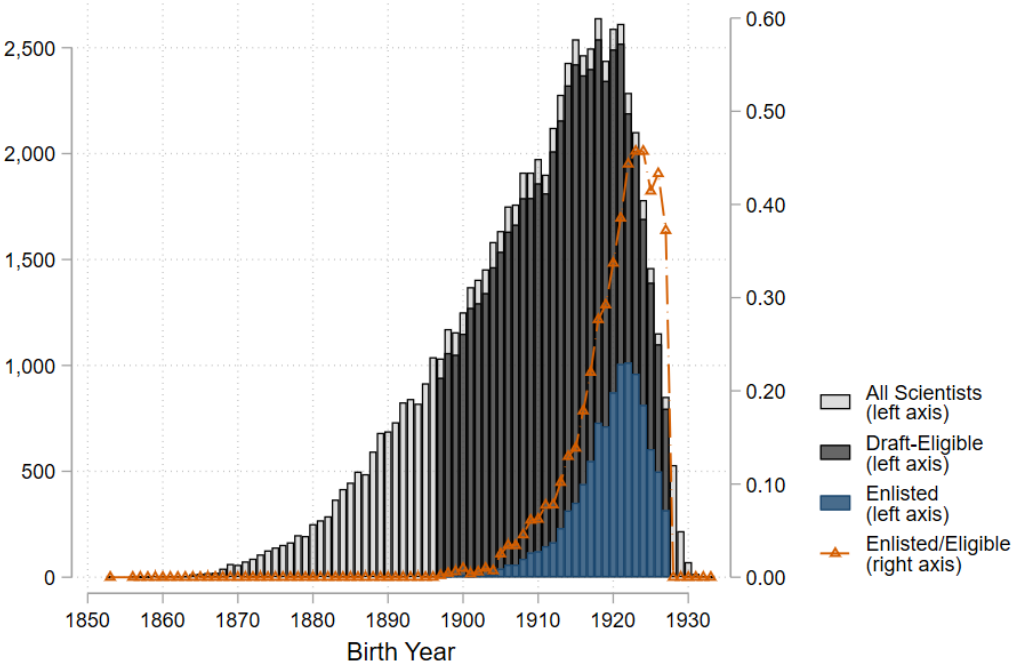
Note: Model-implied talent thresholds and additional female inventors for different distortion scenarios. The relative change in female inventors is calculated as the change in the share of women who enter patenting in the physical science given the respective minimum talent threshold.

Figure 1: Share of Women among Inventors Born between 1921-1980



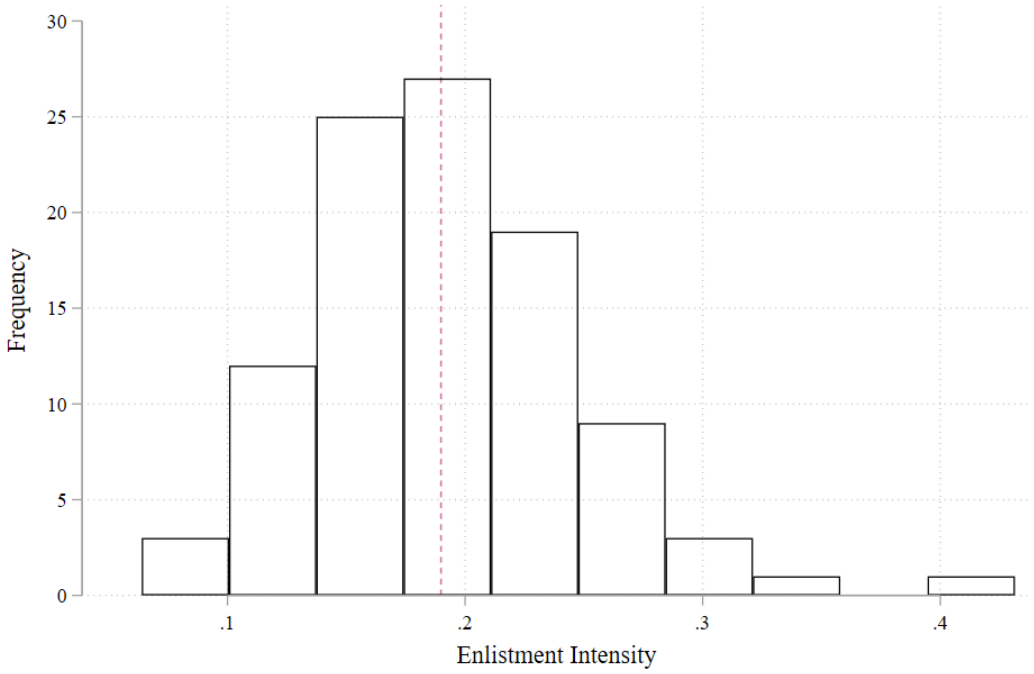
Note: Women inventor share timeline combining data from MoS (1956) and Bell et al. (2019). We define the women inventor share in the MoS (1956) as the percentage of women among all scientists born in a given birth year cohort who ever file a patent in their career.

Figure 2: Enlistment by Birth Cohort



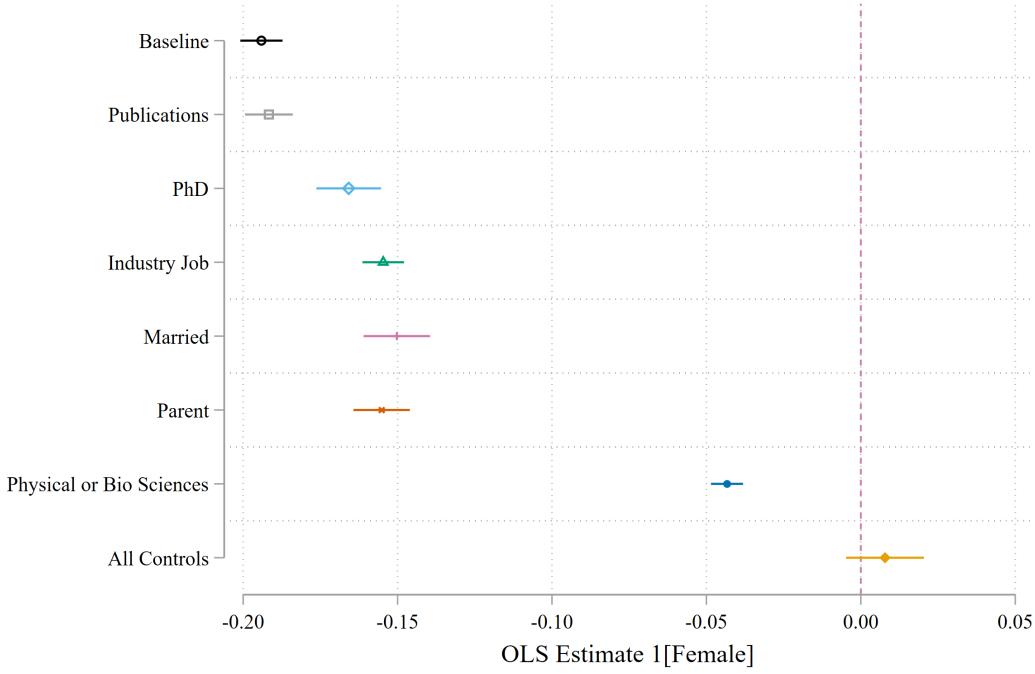
Note: Scientists matched with enlistment records from Chen and Moser (2024). Draft-eligible scientists defined as men who are born between 1897 and 1927.

Figure 3: Enlistment across Fields



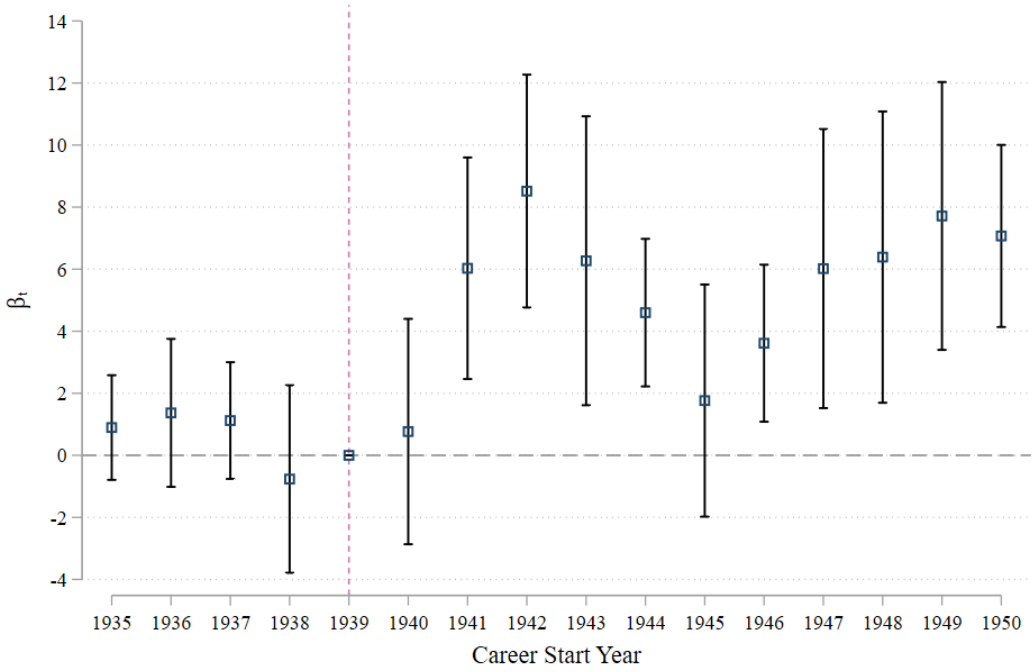
Note: Share of draft-eligible scientists who enlisted by k-means field. Scientists matched with enlistment records from Chen and Moser (2024). Draft-eligible scientists defined as men who are born between 1897 and 1927.

Figure 4: OLS Decomposition of the Innovation Gender Gap



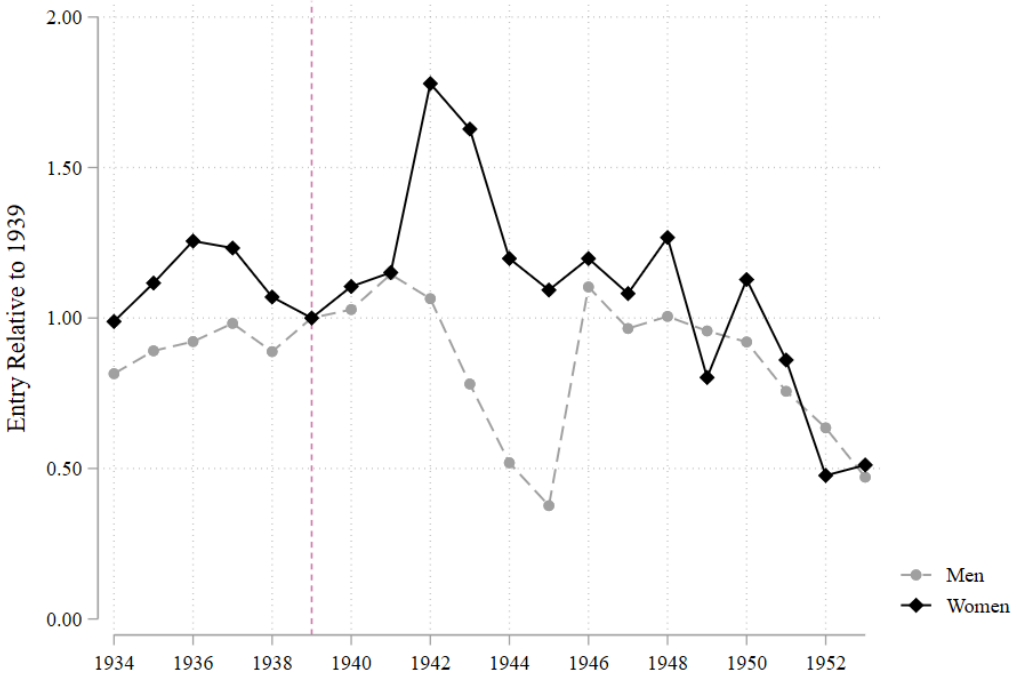
Note: OLS estimate for $\mathbb{1}[Female]_i$ across specifications reported in Table 2, point estimates with 95% confidence intervals. All specifications contain birth year and birth state fixed effects.

Figure 5: Event Study Estimates: Continuous Treatment



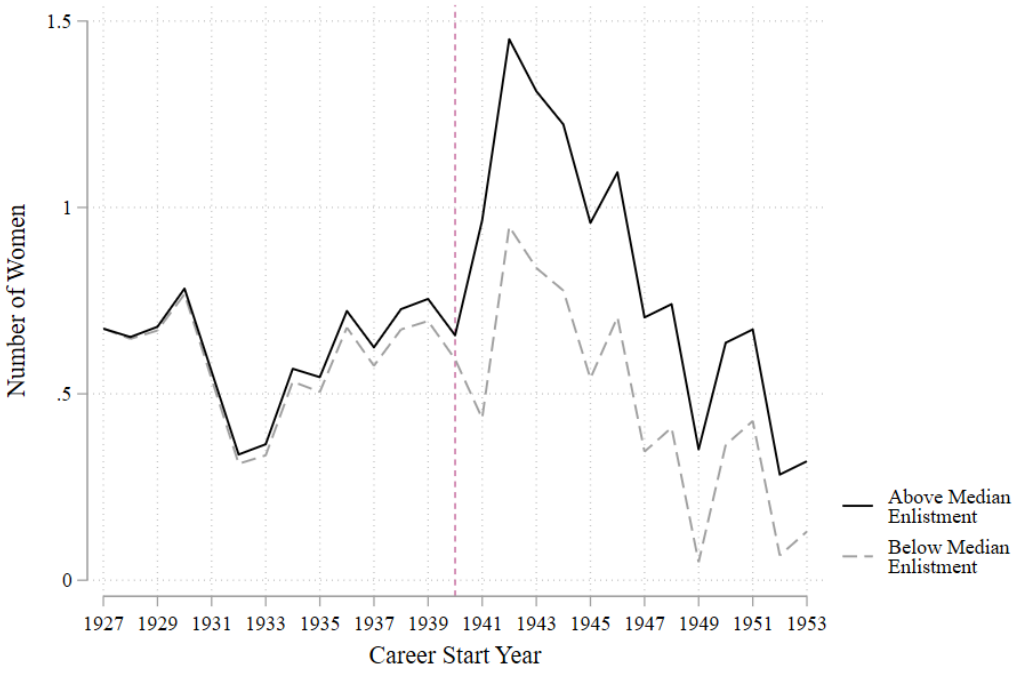
Note: Point estimates and 95% confidence intervals for OLS estimates of $Y_{kt} = \alpha_k + \phi_t + \sum_{t,t\neq 1939}^{T} \left(\mathbb{1}[T = t] \times D_k\right)\beta_t + \varepsilon_{kt}$ on entry cohorts across k-means fields.

Figure 6: Male and Female Entry into Science



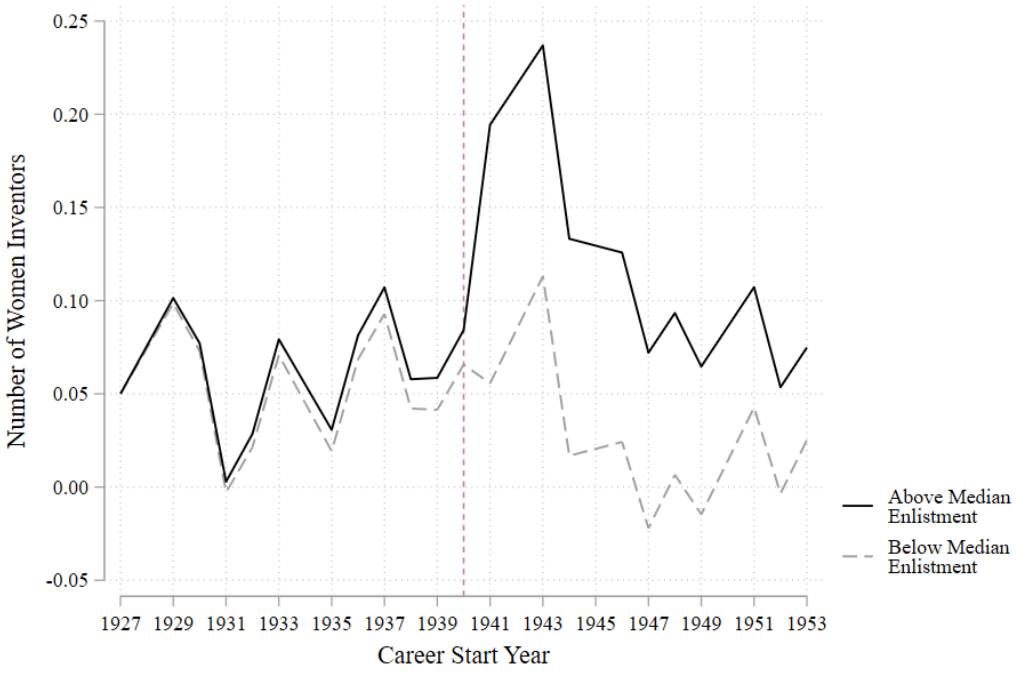
Note: Indexed entry cohort size for male and female entrants in the MoS (1956). Normalized to 1 in 1939.

Figure 7: Entry of Female Scientists into the Physical Sciences



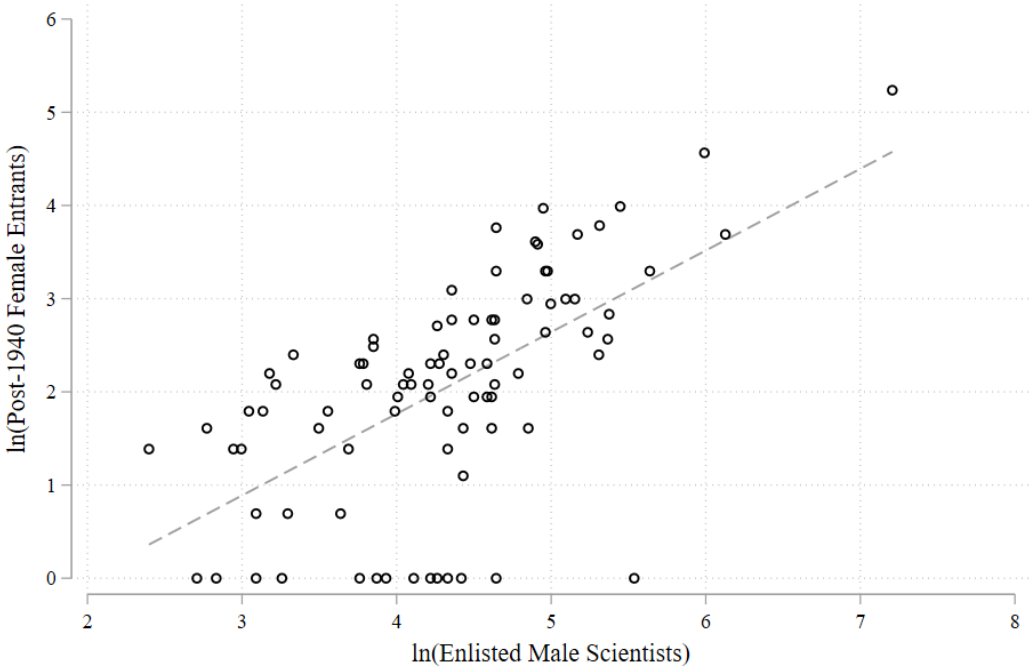
Note: Mean entry of female scientists into *k*-means fields in the physical sciences over time. Smoothed and centered averages across *k*-means fields.

Figure 8: Entry of Female Inventors into the Physical Sciences



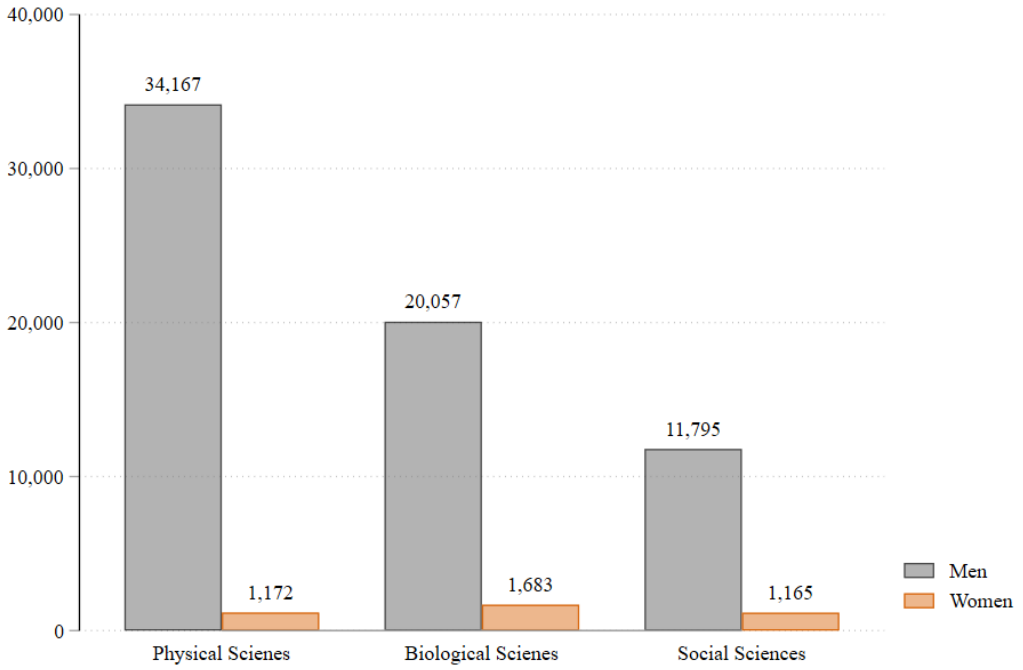
Note: Mean entry of female inventors into *k*-means fields in the physical sciences over time. Smoothed and centered averages across *k*-means fields.

Figure 9: Female Entry After the WWII Draft



Note: Natural log of post-1940 female entry against natural log of male army enlistments across k-means fields. The correlation between the logged measures is $\rho = 0.61$. The correlation between the underlying raw measures is $\rho = 0.89$. The WWII draft began in 1940. Sample limited to draft-eligible birth cohorts (1897-1927).

Figure 10: Male and Female Scientists in the Physical, Biological, and Social Sciences



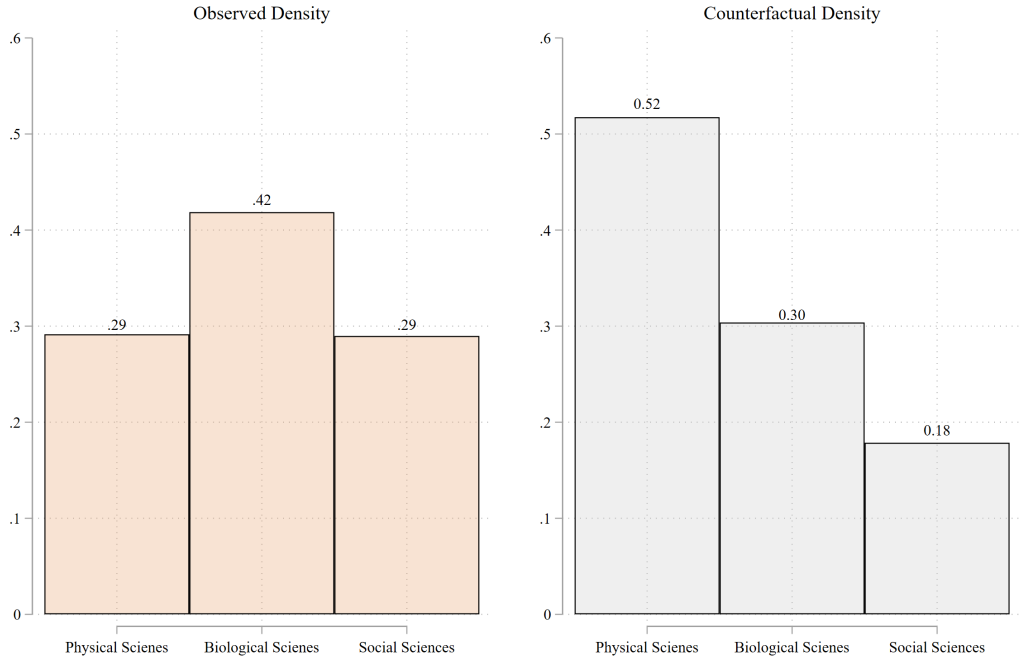
Note: Count of male and female scientists across the physical, biological, and social sciences in the MoS (1956). Women make up 3.3% of scientists in the physical sciences, 7.7% in the biological sciences, and 9.0% in the social sciences.

Figure 11: Share of Female Scientists with a PhD by Size of Female Entry Cohort



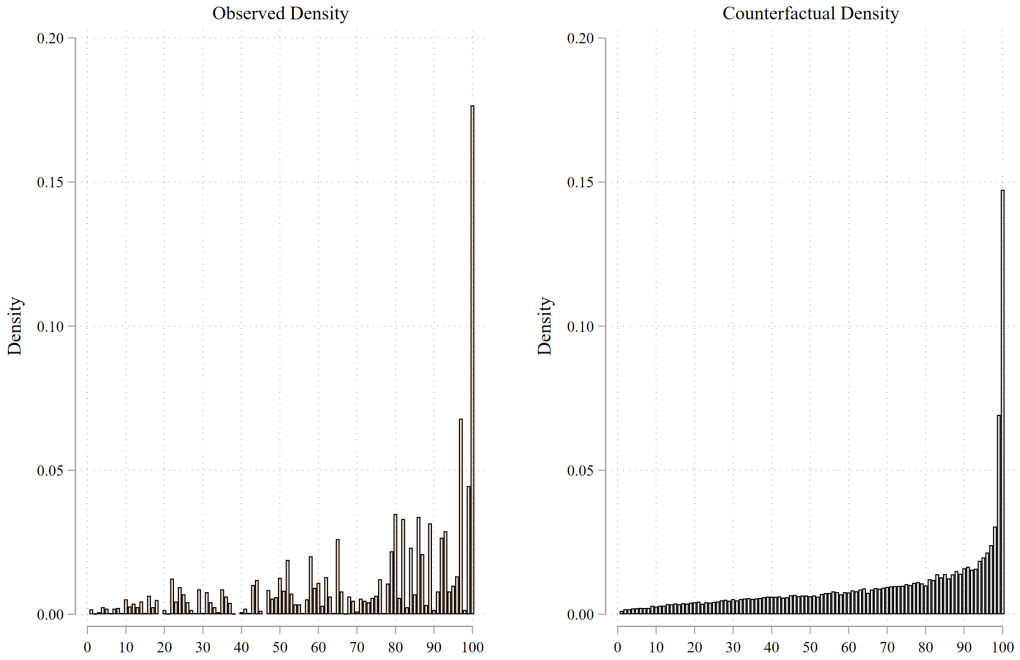
Note: Share of female scientists with a PhD by entry cohort against the size of the female entry cohort (linear fit: -0.0010**). WWII entry cohorts (1940-46) highlighted.

Figure 12: Observed and Counterfactual Densities: Sciences



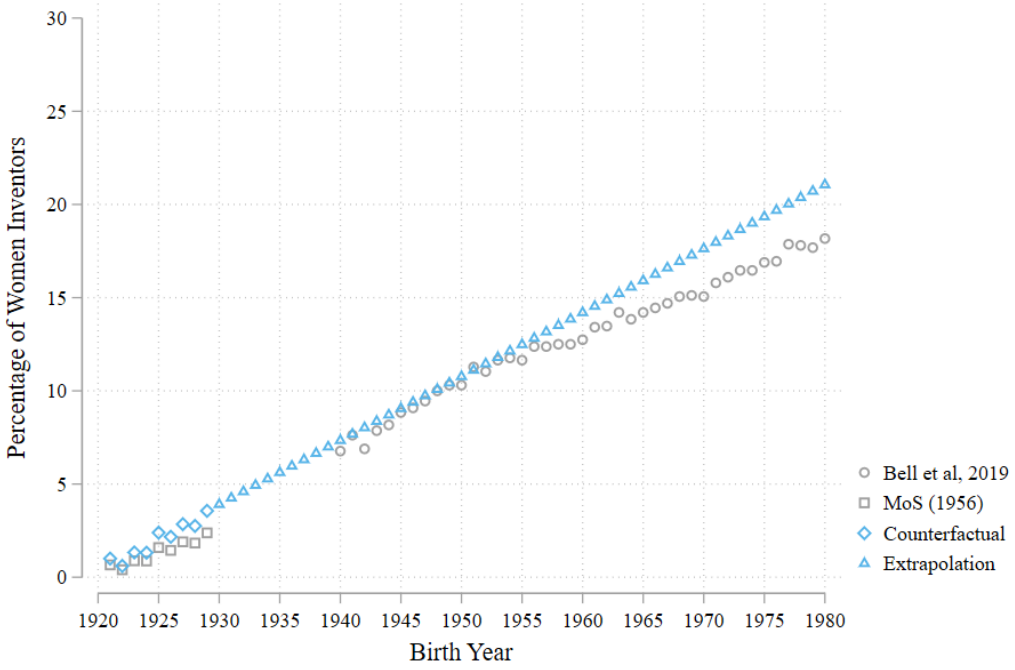
Note: Left: Observed distribution of women in the MoS (1956) across physical, biological, and social sciences. Right: Distribution of men in the MoS (1956) across physical, biological, and social sciences. Reweighting counterfactuals use the distribution of men across sciences as a counterfactual for the distribution of women.

Figure 13: Observed and Counterfactual Densities: k-Means Fields



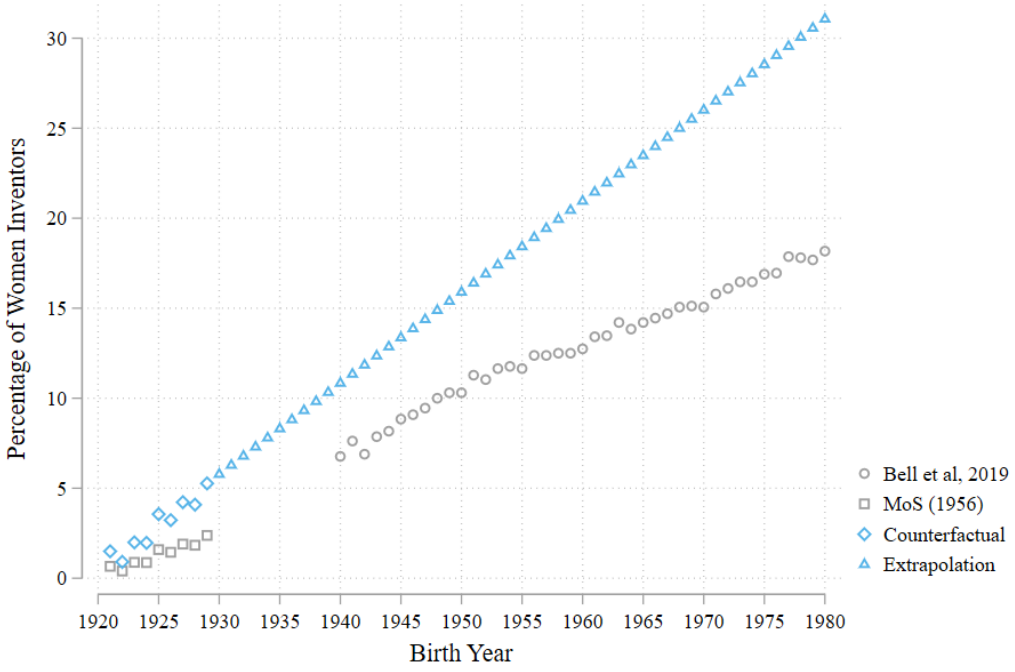
Note: Left: Observed distribution of women in the MoS (1956) across fields. Right: Distribution of men in the MoS (1956) across fields. Reweighting counterfactuals use the distribution of men across sciences as a counterfactual for the distribution of women. Fields ordered by the total number of scientists active in each field.

Figure 14: Counterfactual (1): Equity in the Allocation of Talent across Sciences



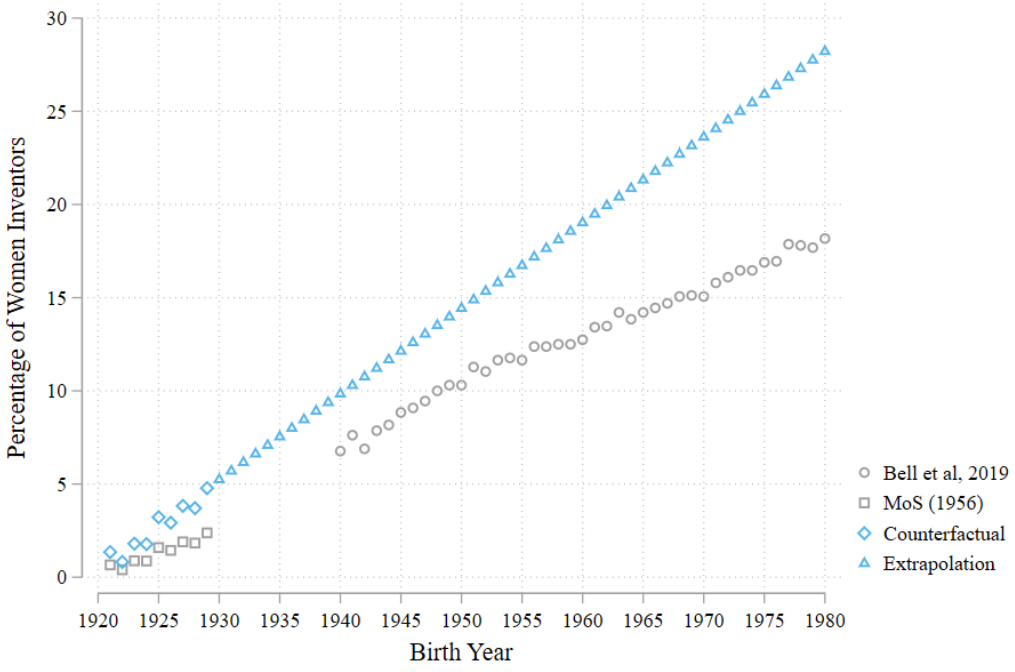
Note: Reweighting female scientists across sciences and keeping conditional patent rates constant.

Figure 15: Counterfactual (2): Equity in the Allocation of Talent across Fields



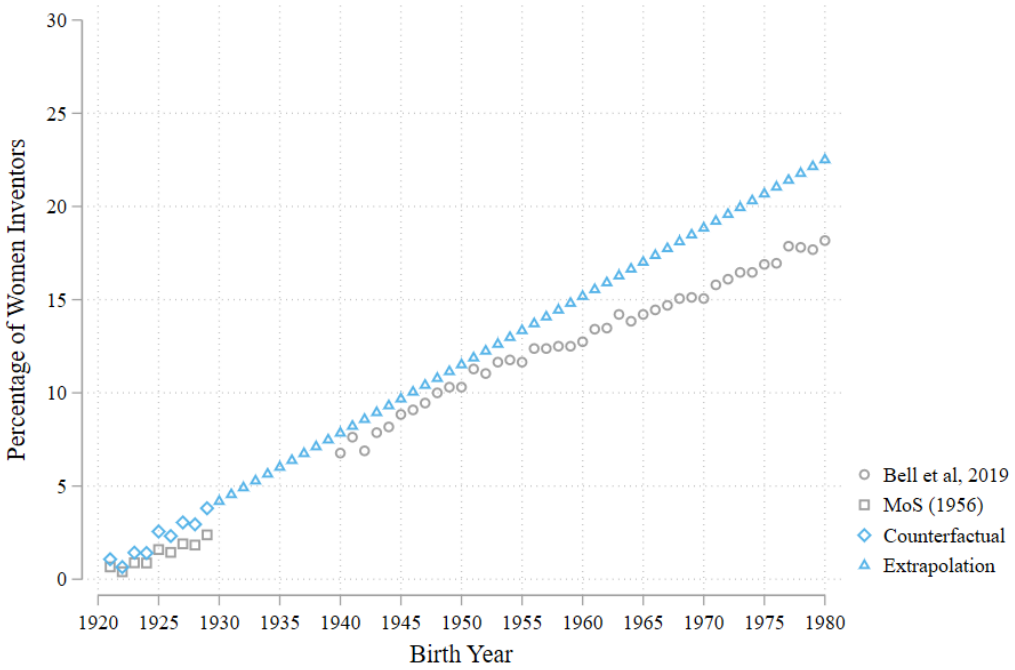
Note: Reweighting female scientists across k-means fields and keeping conditional patent rates constant.

Figure 16: Counterfactual (3): Equity in the Allocation of Talent across Sciences and in Patenting across Fields



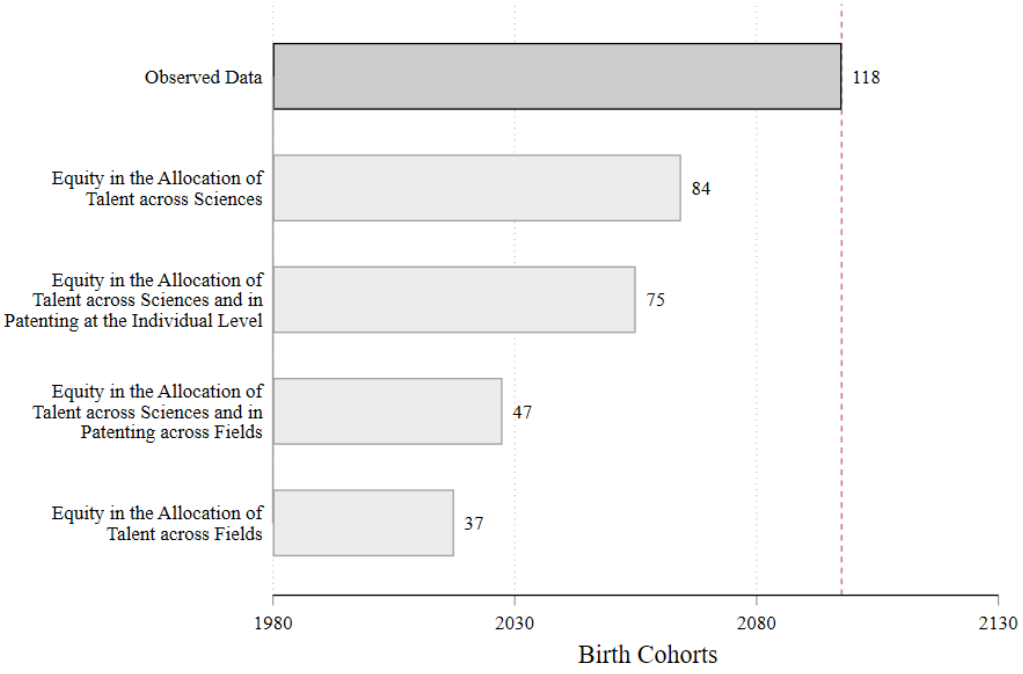
Note: Reweighting female scientists across sciences and adjusting conditional patent rates by field-level IV estimates.

Figure 17: Counterfactual (4): Equity in the Allocation of Talent across Sciences and in Patenting at the Individual Level



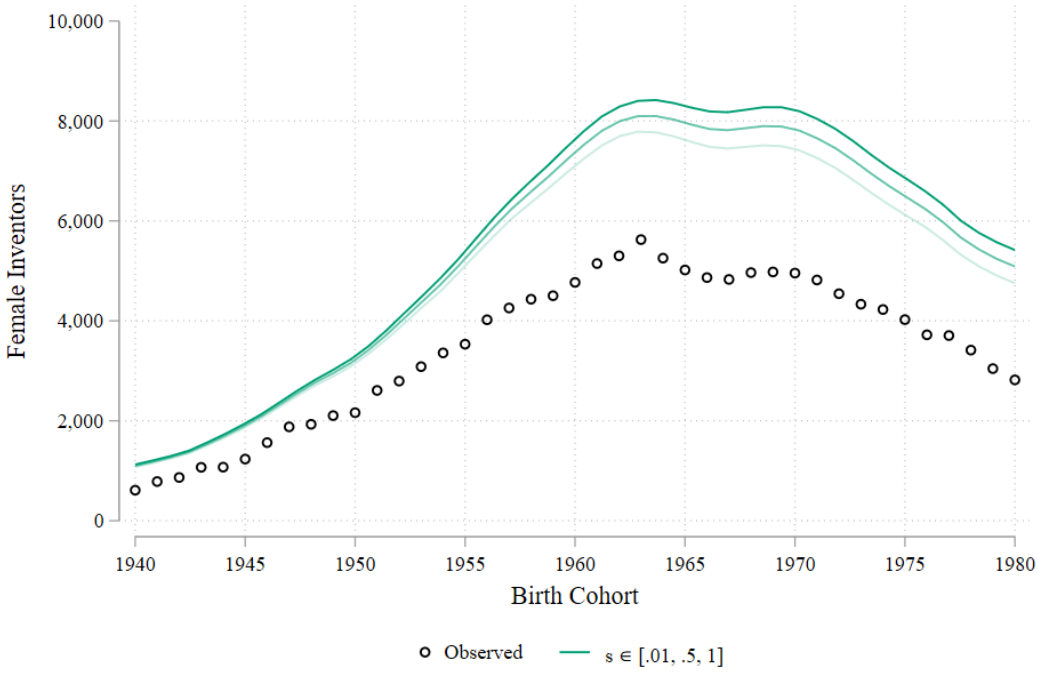
Note: Reweighting female scientists across sciences and adjusting conditional patent rates by individual-level IV estimates.

Figure 18: Counterfactual Timelines to Gender Parity



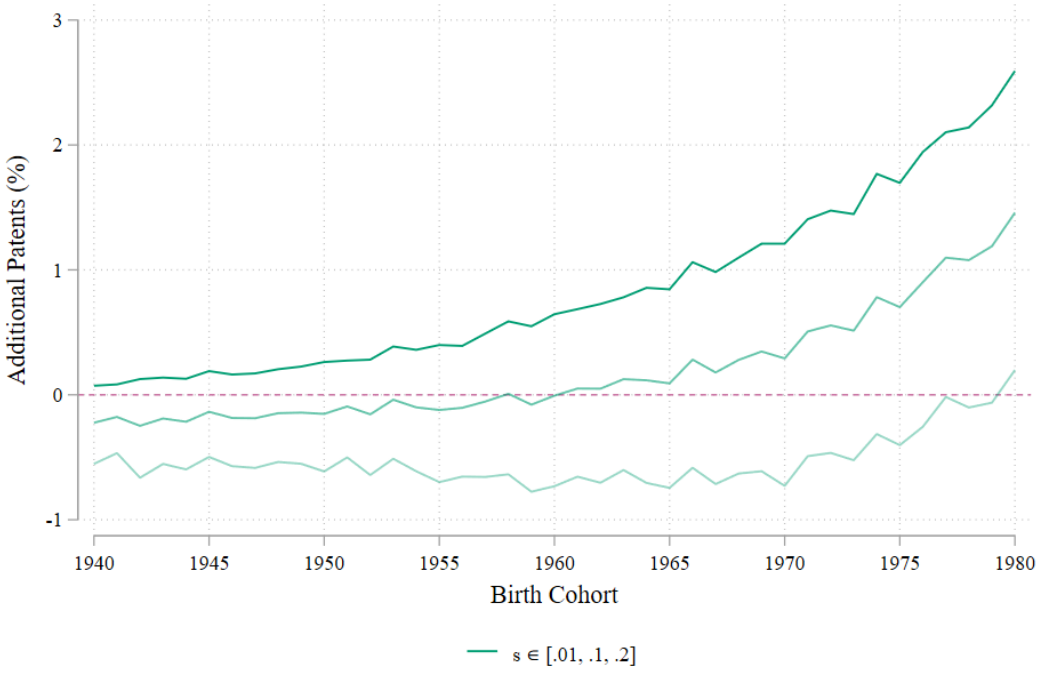
Note: Counterfactual timelines to gender parity in innovation. Observed data from Bell et al. (2019).

Figure 19: Additional Female Inventors



Note: Additional female inventors by birth cohort and substitution rate *s*.

Figure 20: Additional Inventions



Note: Additional inventions by birth cohort and substitution rate *s*.

Figure 21: Game Structure for Three Stage Roy (1951) Model

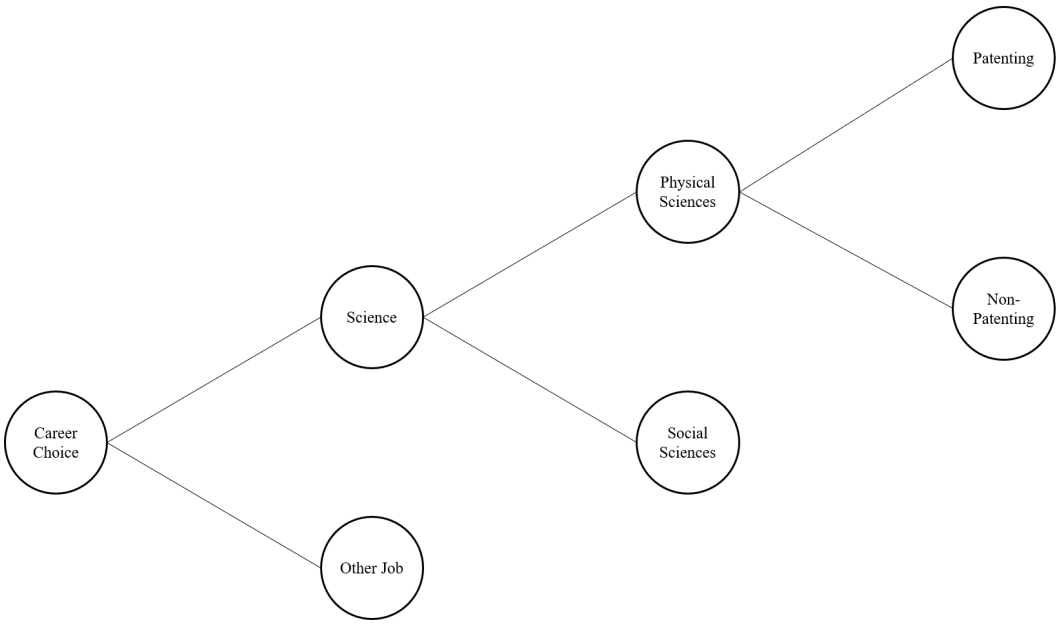
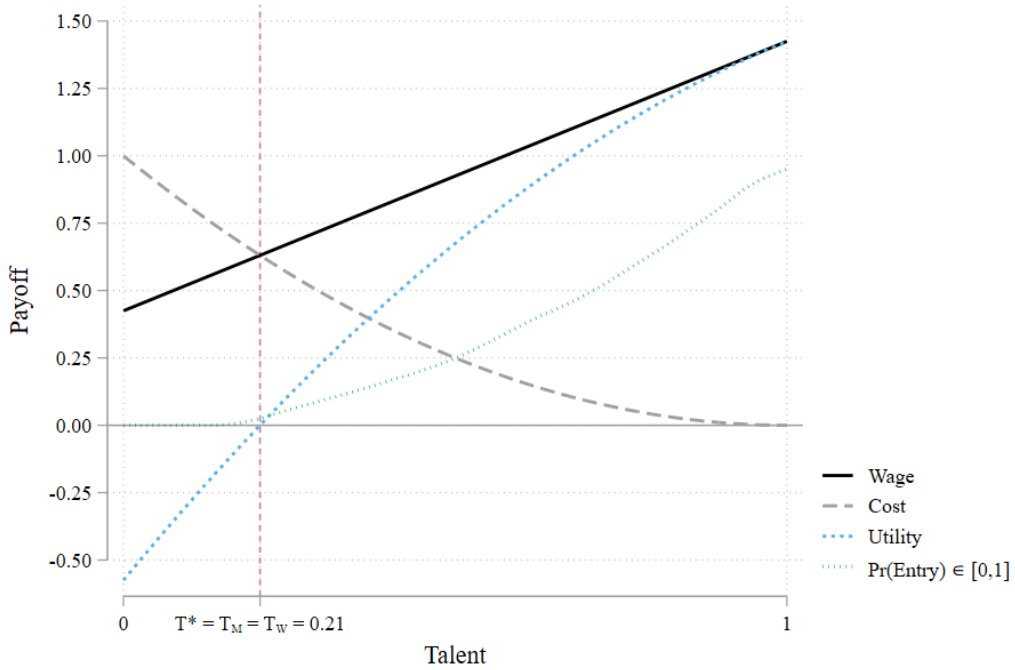
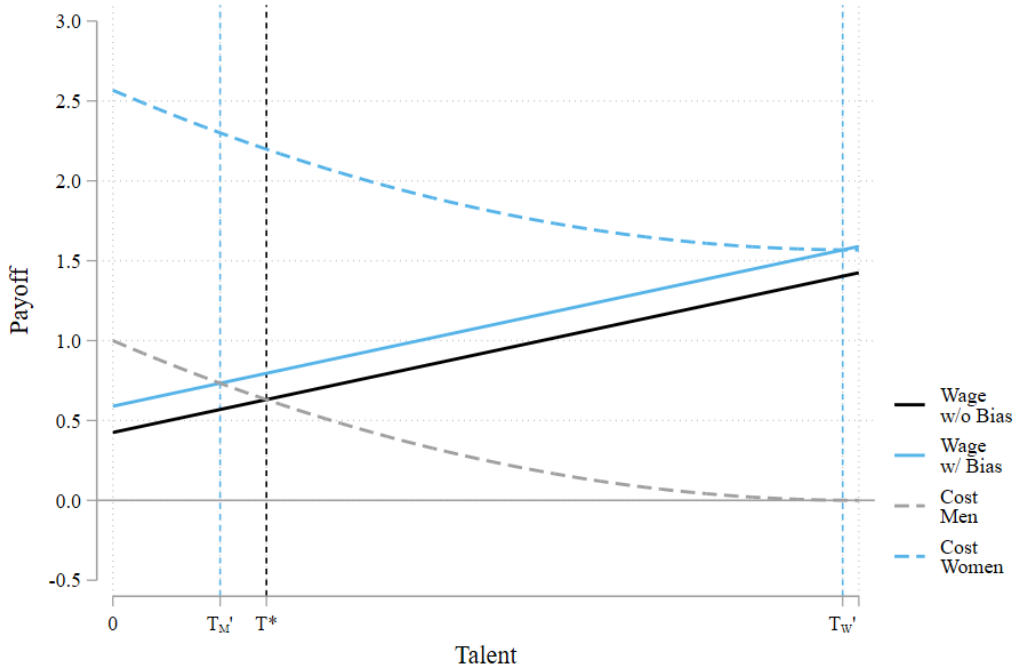


Figure 22: Symmetric Model Equilibrium



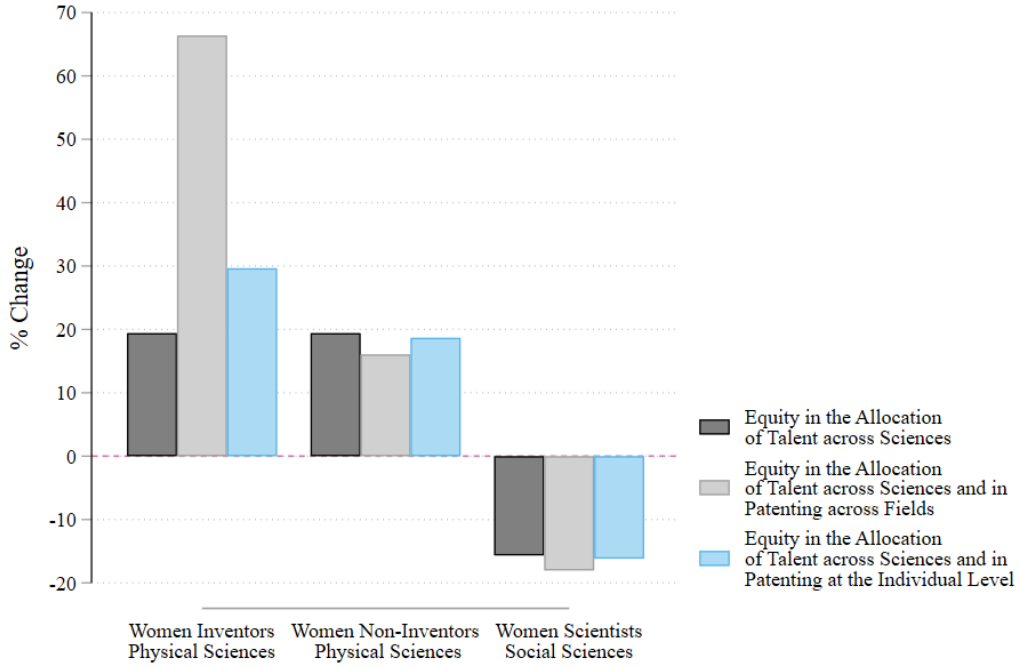
Note: T^* marks the symmetrical equilibrium for the model without distortions, in which the minimum talent thresholds for male and female entrants are identical. T^* marks the location where the cost and wage curves intersect and utility is equal to zero. Entry probabilities obtained from simulating the model with $N = 200,000$ observations. The entry probability is strictly less than one because at any talent level a share of individuals will still be more talented in another *f*.

Figure 23: Model Equilibrium with Entry Bias



Note: T^* marks the symmetrical equilibrium for the model without distortions, in which the minimum talent thresholds for male and female entrants are identical. T'_m and T'_w denote the minimum talent thresholds for male, respectively female entrants that result in equilibrium in the model with entry bias in which only 6% of scientists are women. $T^* = 0.21$, $T'_m = 0.14$, $T'_w = 0.98$.

Figure 24: Counterfactual Increase in Female Inventors – Three-Stage Roy Model



Note: Share of women who become inventors, are active in the physical sciences, and in the social sciences by counterfactual scenarios.

A. Additional Tables

Table A.1: OLS Predictions of Pre-1940 Field Characteristics using Enlistment

Y =	(1) Enlistment Share	(2) Enlistment Share	(3) Enlistment Share
Pre-1940 Women Scientists	0.000		
	(0.000)		
Pre-1940 Women Inventors		-0.005	
		(0.003)	
Pre-1940 Field Size			-0.000
			(0.000)
Mean Y	0.193	0.193	0.193
Observations	100	100	100
R-squared	0.006	0.009	0.013

Note: Robust errors in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.1

Table A.2: Individual-level IV Estimates Controlling for Education and Employment

Y =	(1) Inventor	(2) Inventor	(3) Physical Sciences	(4) Biological Sciences	(5) Inventor
	OLS	Reduced Form	First Stage	First Stage	IV
Physical Sciences (1/0)	0.112***				0.114***
	(0.023)				(0.024)
Bio. Sciences (1/0)	0.033***				0.034***
	(0.006)				(0.006)
Phys. Pull		-9.721***	-86.128***	3.425**	
		(2.076)	(1.710)	(1.522)	
Bio. Pull		-2.709***	0.802	-81.837***	
		(0.468)	(0.646)	(2.073)	
PhD (1/0)	0.008	0.007	-0.009	0.015*	0.008
	(0.012)	(0.012)	(0.006)	(0.008)	(0.012)
Industry Job	0.054***	0.054***	0.004	0.001	0.054***
	(0.020)	(0.020)	(0.011)	(0.011)	(0.020)
Year FE	✓	✓	✓	✓	✓
Mean Y	0.045	0.045	0.288	0.394	0.045
Observations	1,951	1,951	1,951	1,951	1,951
Cragg-Donald F					8552.021
Kleibergen-Paap F					1047.170

Note: IV regressions estimated with 2SLS. Inventor is an indicator for whether a scientist has at least one patent. Physical and Biological Sciences are indicator variables for scientists in the respective sciences. PhD is an indicator variable for scientists with a doctoral degree. Industry is an indicator variable for scientists in industry jobs. Sample limited to female scientists. Robust errors clustered by field in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.10

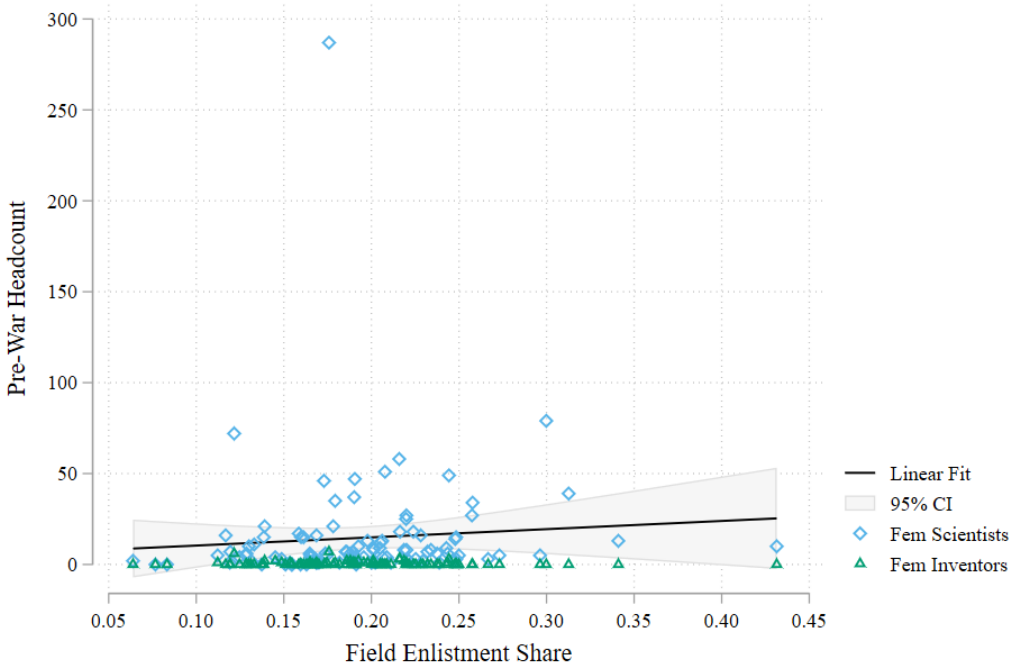
Table A.3: Individual-level IV Estimates with Controls and Overidentifying Restrictions Test

Y =	(1) Inventor	(2) Inventor	(3) Physical Sciences	(4) Biological Sciences	(5) Inventor
	OLS	Reduced Form	First Stage	First Stage	IV
Physical Sciences (1/0)	0.112***				0.114***
	(0.023)				(0.024)
Bio. Sciences (1/0)	0.033***				0.034***
	(0.006)				(0.006)
Phys. Pull		-9.903***	-84.213***	4.899***	
		(2.275)	(2.035)	(1.527)	
Bio. Pull		-3.122***	1.613*	-81.593***	
		(0.642)	(0.871)	(2.008)	
Phys. Pull × Enlistment		-0.839	-3.625***	-4.127***	
		(0.653)	(1.110)	(1.262)	
Bio. Pull × Enlistment		0.805	1.841**	2.524***	
		(0.508)	(0.740)	(0.817)	
PhD (1/0)	0.008	0.007	-0.006	0.017**	0.008
	(0.012)	(0.012)	(0.006)	(0.008)	(0.012)
Industry Job	0.054***	0.054***	0.005	0.002	0.054***
	(0.020)	(0.020)	(0.010)	(0.010)	(0.020)
Year FE	✓	✓	✓	✓	✓
Mean Y	0.045	0.045	0.288	0.394	0.045
Observations	1,951	1,951	1,951	1,951	1,951
Cragg-Donald F					4648.638
Kleibergen-Paap F					587.703
Hansen J					2.002
Hansen J (p-value)					0.368

Note: IV regressions estimated with 2SLS. Inventor is an indicator for whether a scientist has at least one patent. Physical and Biological Sciences are indicator variables for scientists in the respective sciences. PhD is an indicator variable for scientists with a doctoral degree. Industry is an indicator variable for scientists in industry jobs. Sample limited to female scientists. Enlistment is the share of enlisted draft-eligible scientists in the same k-means field. Sample limited to female scientists. Robust errors clustered by field in parentheses.
*** p < 0.01 ** p < 0.05 * p < 0.10

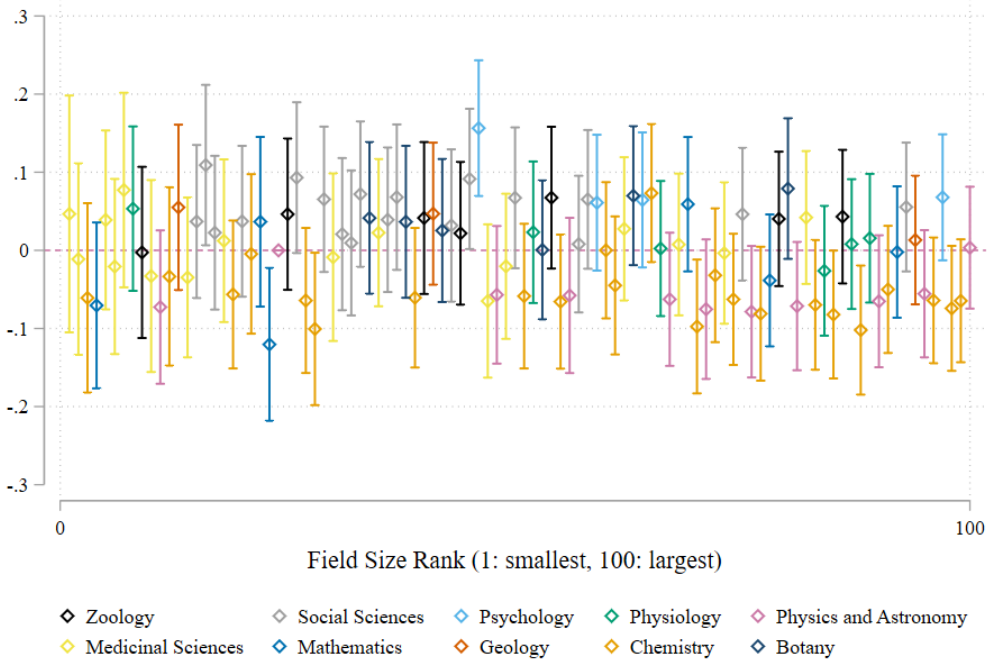
B. Additional Figures

Figure B1: Correlation between Enlistment Share and Pre-1940 Female Entry



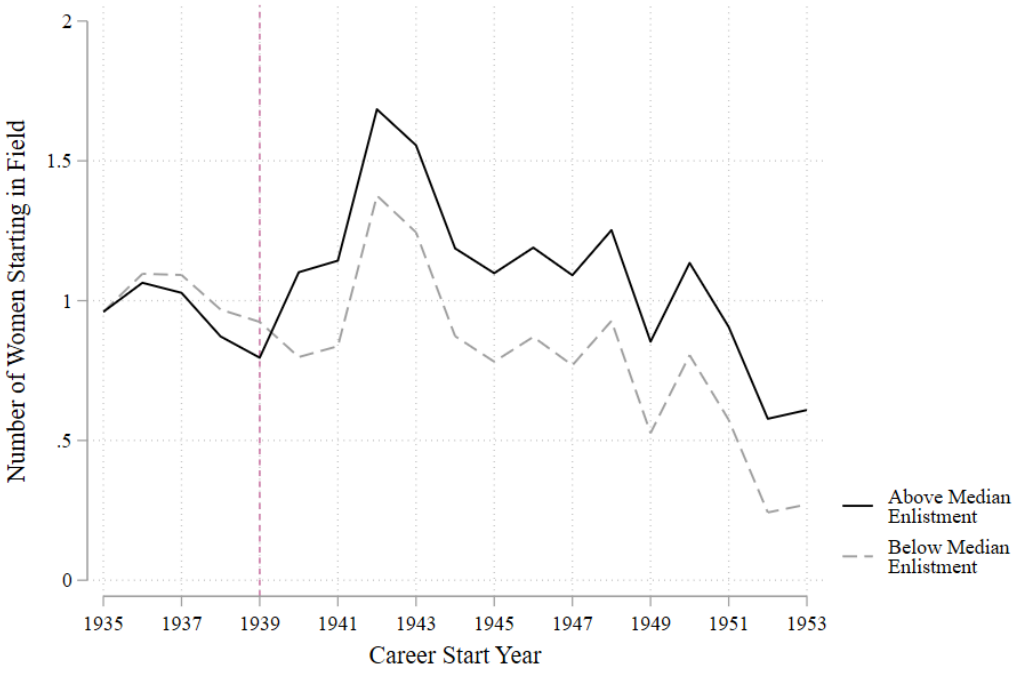
Note: Y-axis: Count of female scientists and female inventors per field in the MoS (1956) before the start of the WWII draft in 1940. X-axis: Share of male scientists per field who enlisted in the army during WWII. Linear fit: predicted values for the number of female scientists in a field from a linear regression of female scientists on field-level enlistment rates. The correlation coefficient between female scientists and enlistment is 0.12, the correlation coefficient between female inventors and enlistment is -0.02.

Figure B2: OLS Predicting Enlistment with Field Dummies



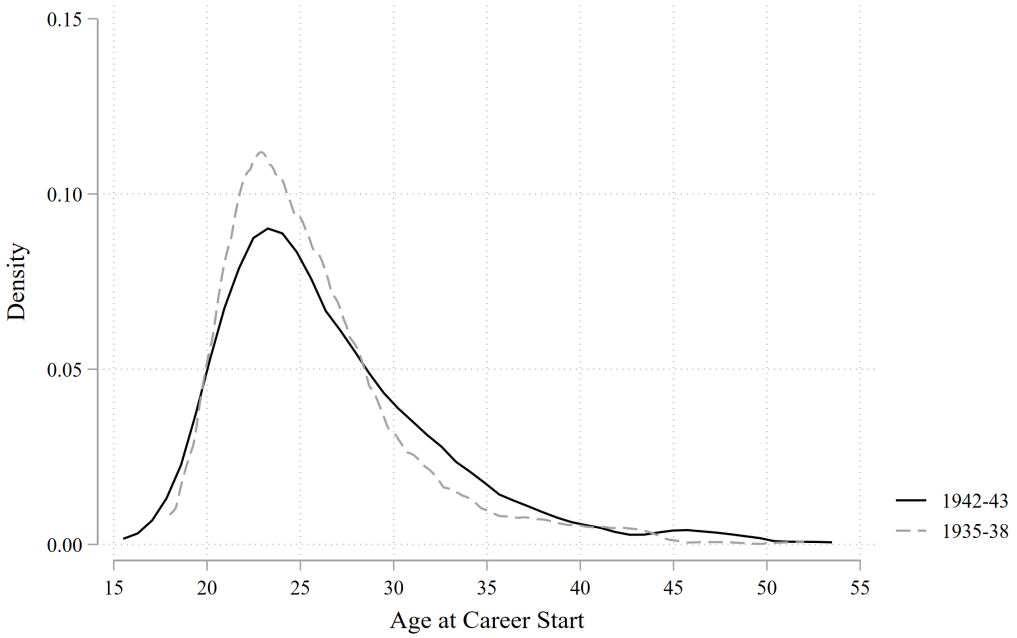
Note: OLS coefficients for k-means field fixed effects with 95% confidence intervals for a prediction of $Pr(Enlisted)_i$ for male scientists born between 1910-27. Enlistment is a binary variable. Specification further includes control variables for marital status, parenthood, birth year, birth state, birth year-by-birth state fixed effects. Out of 100 estimated coefficients, 6 are significant at 5% and one is significant at 1%.

Figure B3: Mean Female Entry at Field-Level



Note: Smoothed and centered average female entry in fields with above and below median enlistment among male scientists.

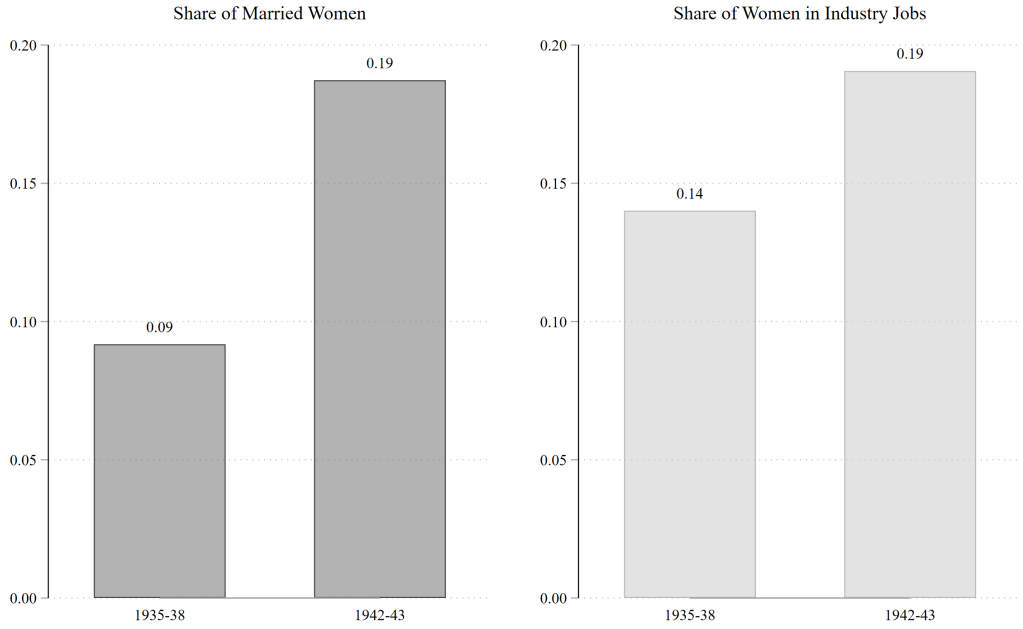
Figure B4: Age Distribution of Female Entrants, 1935-38 and 1942-43



N = 713

Note: Kernel density estimates of women's age at entry for entry cohorts during (1942-43) and before the start of the WWII draft (1935-38) in the MoS (1956). A two-sample t-test allowing for unequal variances indicates that average age at entry is significantly higher in the 1942-43 cohorts with a p-value of 0.019.

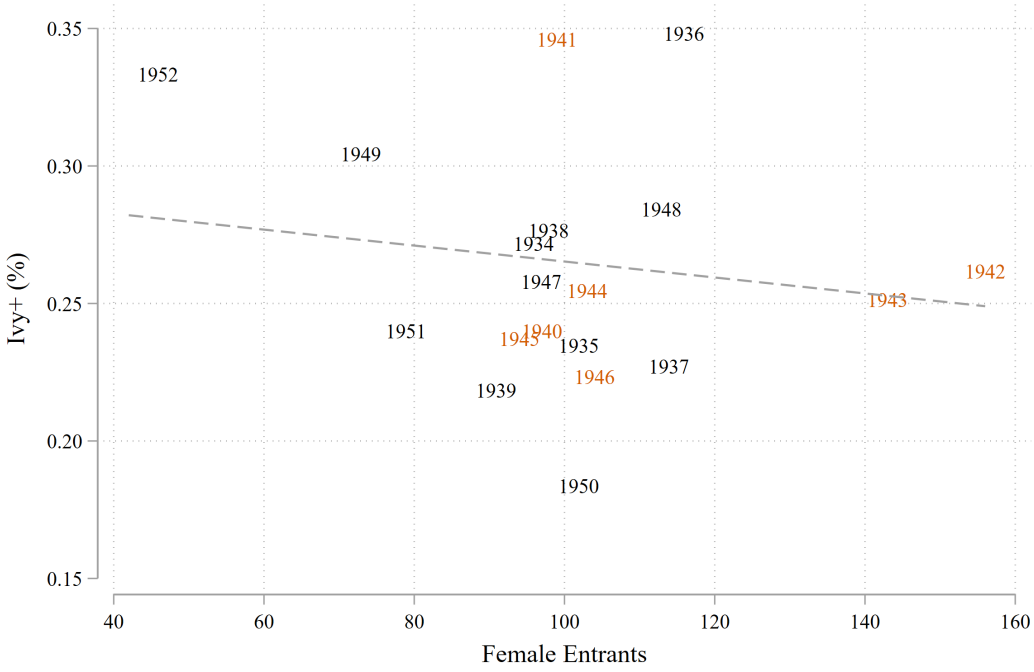
Figure B5: Share of Married Women and Women in Industry Jobs, 1935-38 and 1942-43



N = 713

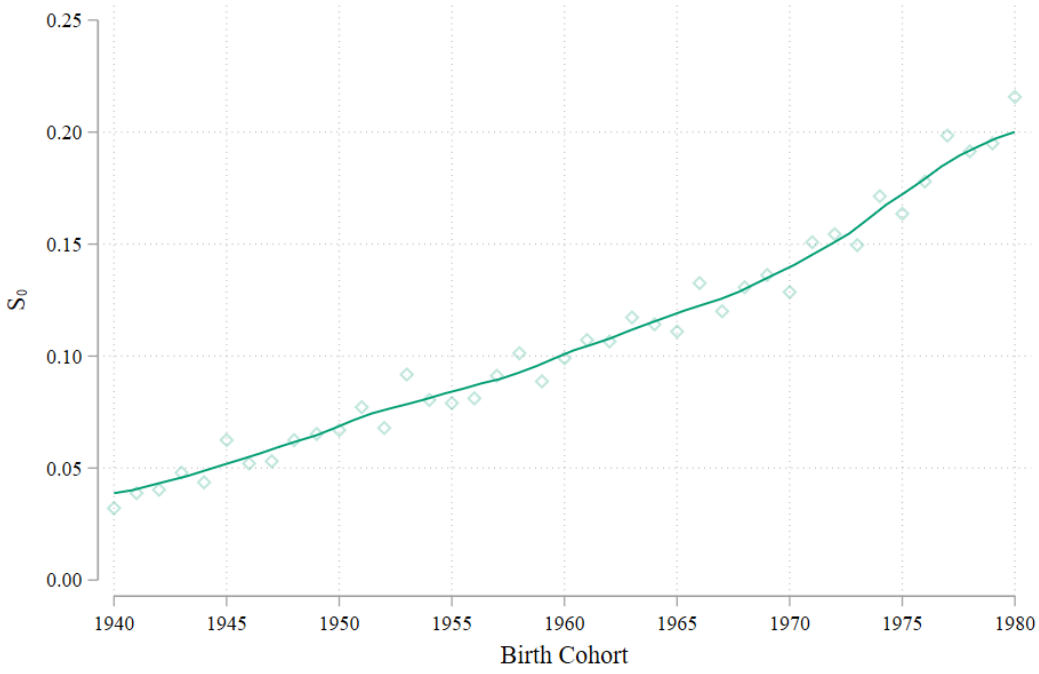
Note: Left: Share of married women among women entering science before (1935-38) and during WWII (1942-43). Marriage measured in year of entry. Right: Share of women working in industry jobs among women entering science before (1935-38) and during WWII (1942-43). Two-sample t-tests allowing for unequal variances indicates that the share of women married at entry is significantly higher in the 1942-43 cohorts with a p-value of 0.000 and that the share of women working in industry is significantly higher in the 1942-43 cohorts with a p-value of 0.038.

Figure B6: Share of Scientists with an Ivy+ Degree by Size of Female Entry Cohort



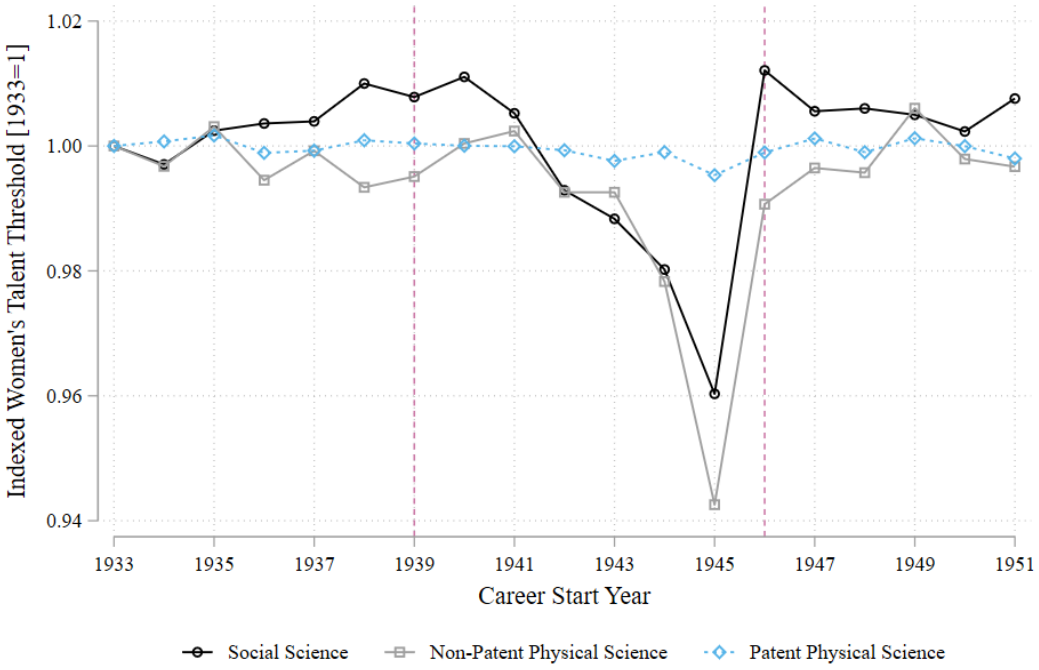
Note: Share of female scientists with an Ivy+ degree by entry cohort against the size of the female entry cohort (linear fit: -0.0003). WWII entry cohorts (1940-46) highlighted.

Figure B7: Breakeven Substitution Rate



Note: Substitution rate s at which aggregate invention effect of counterfactual entry becomes positive.

Figure B8: Three-Stage Roy Model: Women's Talent Thresholds during WWII



Note: Model-implied talent thresholds for women entering the social sciences, non-patenting physical sciences, and patenting physical sciences. Talent thresholds normalized to 1 in 1933. Talent thresholds derived from calibrating model separately for every scientist entry cohort.

C. Model Appendix

We develop a three-stage Roy (1951) model to investigate gender bias in three stages: entry into science, field choice, and patenting (Figure 21). In the first stage, individuals i choose to work in science or another job; the utility of working in another job is normalized to zero. In the second stage, scientists choose between research fields in the physical and social sciences. In the third stage, individuals who specialized in STEM (the physical sciences) can choose to patent, while social scientists (e.g., in psychology or anthropology) do not patent.

Let $\phi_{f,i}$ denote individual i's talent. The index f takes three values, S, P, or N, where S denotes an individual's talent for the social sciences (and not patenting), while P denotes an individual's talent for patenting in the physical sciences, and N denotes the individual's talent for non-patent research in the physical sciences. Each individual is endowed with a vector $\Phi_i = \{\phi_{S,i}, \phi_{P,i}\, \phi_{N,i}\}$ of i.i.d. standard uniform talents across S, P, and N.

The distribution of talent for f is identical for women and men. An individual's talent in f determines both the wage they earn from specialising in f, and the cost of specialization c. After individuals see their talent draws, they choose whether to become a scientist, which field to enter, and whether to patent by maximizing their utility:

$$\max_f u_i = \max[w(\phi_{f,i}) - c(\phi_{f,i}), 0] \tag{26}$$

where $w(\phi_{f,i})$ is the wage function and c captures the cost of individual i working in working in f. The cost of specializing in f, $c(\phi_{f,i})$, for individual i is convex in their talent, so that more talented scientists face lower costs of entry: $c(\phi_{f,i}) = (1-\phi_{f,i})^2$. Wages increase additively in individual i's own talent in f and in the average talent of all people in f, but decrease as the field becomes more crowded and a larger share of individuals are active in f.

$$w(\phi_{f,i}) = \phi_{f,i} + \theta_f \mu_w^f + (1-\theta_f)\mu_m^f - \omega(1-F_w(\bar{f}_w)) - (1-\omega)(1-F_m(\bar{f}_m)) \tag{27}$$

μ_w^f and μ_m^f capture the average talent of female and male researchers in f. θ_f is the share of women among scientists in f.[30] $(1-F_w(\bar{f}_w))$ and $(1-F_m(\bar{f}_m))$ denote the share of women and men, respectively, who become scientists in f, given by the multivariate cumulative density function of Φ_i at $\bar{f}$, the minimum talent required to enter f. Intuitively, the share of entrants is given by the population share of individuals who have drawn their highest talent in f and whose talent in f exceeds the talent threshold in f: these individuals have a comparative advantage in f. The share of entrants is determined jointly by the share of women who are above the talent threshold for women and by the share of men who are above the talent

[30] $\theta_f = \frac{(1-F_w(\bar{f}_w))}{(1-F_w(\bar{f}_w))+(1-F_m(\bar{f}_m))}$

threshold for men $\omega(1-F_w(\bar{f}_w))+(1-\omega)(1-F_m(\bar{f}_m))$. While the population share of women ω is constant, the share of women in f, θ_f, varies with entry decisions.

We specify an individual's wage equation in f as follows. Consider the example f=P, where the share of entrants given a minimum talent threshold $\bar{P}$ is given by (gender subscripts omitted):

$$1-F(\bar{P}) = \int_{\bar{P}_w}^{1}\int_{0}^{\phi_P}\int_{0}^{\phi_P} f(\phi_P,\phi_N,\phi_S)\, d\phi_P d\phi_N d\phi_S \tag{28}$$

$$= \int_{\bar{P}_w}^{1}\int_{0}^{\phi_P}\int_{0}^{\phi_P} 1\, d\phi_N d\phi_S d\phi_P \tag{29}$$

$$= \int_{\bar{P}_w}^{1}\int_{0}^{\phi_P} \phi_N\Big|_0^{\phi_P} d\phi_S d\phi_P \tag{30}$$

$$= \int_{\bar{P}_w}^{1}\int_{0}^{\phi_P} \phi_P\, d\phi_S d\phi_P \tag{31}$$

$$= \int_{\bar{P}_w}^{1} \phi_P\phi_S\Big|_0^{\phi_P} d\phi_P \tag{32}$$

$$= \int_{\bar{P}_w}^{1} \phi_P^2\, d\phi_P \tag{33}$$

$$= \frac{1}{3}\phi_P^3\Big|_{\bar{P}}^{1} \tag{34}$$

$$= \frac{1}{3} - \frac{1}{3}\bar{P}^3 \tag{35}$$

$$= \frac{(1-\bar{P}^3)}{3} \tag{36}$$

Assuming that talent is independently and identically distributed as standard uniform across fields, entry shares for N and S are:

$$1-F(\bar{N}) = \frac{(1-\bar{N}^3)}{3} \tag{37}$$

$$1-F(\bar{S}) = \frac{(1-\bar{S}^3)}{3} \tag{38}$$

These definitions of the multivariate talent distribution make it possible to define the average talent of scientists in f, μ^f (omitting gender subscripts) as the conditional expectation of talent among entrants. For $f = P$, this is given by: $\mu^P = E[\phi_P|\phi_P > \bar{P}, \phi_P > \phi_N, \phi_P > \phi_S]$. Using the definitions of conditional expectations and conditional densities,[31] we can thus define μ^P as:

[31] $E[X|Y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx$ where $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

$$\mu^P = \frac{\int_{\bar{P}_w}^1 \int_0^{\phi_P} \int_0^{\phi_P} \phi_P \, f(\phi_P, \phi_N, \phi_S) \, d\phi_P d\phi_N d\phi_S}{\int_{\bar{P}_w}^1 \int_0^{\phi_P} \int_0^{\phi_P} f(\phi_P, \phi_N, \phi_S) \, d\phi_P d\phi_N d\phi_S} \tag{39}$$

Substituting equations (8) and (11) from above yields:

$$\mu^P = \left(\int_{\bar{P}_w}^1 \phi_P \phi_P^2 \, d\phi_P \right) \times \left(\frac{1-\bar{P}^3}{3} \right)^{-1} \tag{40}$$

$$= \left(\frac{1}{4} \phi_P^4 \Big|_{\bar{P}}^1 \right) \times \left(\frac{1-\bar{P}^3}{3} \right)^{-1} \tag{41}$$

$$= \frac{\frac{1}{4}(1-\bar{P}^4)}{\frac{1}{3}(1-\bar{P}^3)} \tag{42}$$

$$= \frac{3}{4} \left(\bar{P} + \frac{1}{1+\bar{P}+\bar{P}^2} \right) \tag{43}$$

And equivalently:

$$\mu^N = \frac{3}{4} \left(\bar{N} + \frac{1}{1+\bar{N}+\bar{N}^2} \right) \tag{44}$$

$$\mu^S = \frac{3}{4} \left(\bar{S} + \frac{1}{1+\bar{S}+\bar{S}^2} \right) \tag{45}$$

Indifference conditions and model solution

We use the wage equations to solve the model for the minimum talent thresholds $\bar{S}_w$, $\bar{S}_m$, $\bar{P}_w$, $\bar{P}_m$, $\bar{N}_w$, and $\bar{N}_m$ by backward induction. Each threshold denotes the talent level at which a marginal male or female individual is indifferent about entering. As we show below, in equilibrium all individuals with comparative advantage and talent above the threshold enter and no individual with talent below the threshold enter.

We solve the model by considering the minimum talent thresholds as a system of simultaneously determined equations. In equilibrium, individuals at the marginal talent level have to be indifferent between entering f and choosing a career in science over the outside option. Accordingly, at the marginal talent level, the utility of being in any f has to be equal to the utility of being in any other science occupation. For indifference, the utility at the marginal talent level also has to be equal to zero, the normalized outside option. We specify the following system of equations:

$$w(\bar{P}_w) - c(\bar{P}_w) = w(\bar{N}_w) - c(\bar{N}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{N}_m) - c(\bar{N}_m) \tag{46}$$

$$w(\bar{P}_w) - c(\bar{P}_w) = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{47}$$

$$w(\bar{N}_w) - c(\bar{N}_w) = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{N}_m) - c(\bar{N}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{48}$$

$$w(\bar{P}_w) - c(\bar{P}_w) = 0 \qquad w(\bar{P}_m) - c(\bar{P}_m) = 0 \tag{49}$$

$$w(\bar{N}_w) - c(\bar{N}_w) = 0 \qquad w(\bar{N}_m) - c(\bar{N}_m) = 0 \tag{50}$$

$$w(\bar{S}_w) - c(\bar{S}_w) = 0 \qquad w(\bar{S}_m) - c(\bar{S}_m) = 0 \tag{51}$$

In total, there are 12 equations that we solve for 6 unknown thresholds. Six equations, (20) - (22), describe indifference conditions of choosing between fields, and six equations, (23) - (25) describe indifference conditions of choosing between becoming a scientist in a given field and the outside option. We solve the system of equations numerically, resulting in the following thresholds:

$$\{\bar{P}_w, \bar{P}_m\} = \{0.20585, 0.20585\}$$
$$\{\bar{N}_w, \bar{N}_m\} = \{0.20585, 0.20585\}$$
$$\{\bar{S}_w, \bar{S}_m\} = \{0.20585, 0.20585\}$$

These talent thresholds are symmetrical across all f and across genders. Without additional distortions and under the assumption that there are no gender differences in the inherent distribution of talent for scientific work, men and women should distribute evenly and symmetrically across research fields and patenting.

Equilibrium

We solve the model to find a sub-game perfect Nash equilibrium. For the symmetric solution to constitute such an equilibrium, no individual can have individual incentives to deviate from this solution. To show that symmetric talent thresholds are an equilibrium, we introduce two simple propositions. First, everyone whose talent in f is below the threshold level $\bar{f}$ will not enter. Second, above the talent threshold, utility is strictly increasing in talent, so that everyone who is talented enough in f and has comparative advantage always enters.

Proposition 1. *No individual whose talent in f is below the talent threshold $\bar{F}$ enters f.*

Proof. Consider an individual's utility using the simplified wage function for f given symmetric talent threshold values $\bar{F} = \bar{F}_m = \bar{F}_w$:

$$u_i = w(\phi_{f,i}) - (1-\phi_{f,i})^2 = \phi_{f,i} + \frac{(1-\bar{F}^3)}{3} - \frac{3}{4}\left(\bar{F} + \frac{1}{1+\bar{F}+\bar{F}^2}\right) - (1-\phi_{f,i})^2$$

Now, for an individual to enter with $\phi_{f,i} < \bar{F}$, we require:

$$\phi_{f,i} - (1-\phi_{f,i})^2 \geq \frac{3}{4}\left(\bar{F} + \frac{1}{1+\bar{F}+\bar{F}^2}\right) - \frac{(1-\bar{F}^3)}{3}$$

However, following from our definition of $\bar{F}$:

$$\bar{F} - (1-\bar{F})^2 = \frac{3}{4}\left(\bar{F} + \frac{1}{1+\bar{F}+\bar{F}^2}\right) - \frac{(1-\bar{F}^3)}{3}$$

Thus, for any $\phi_{f,i} < \bar{F}$, this equality will not hold and the individual will prefer the outside option to entering f. □

Proposition 2. *The utility of entering f is strictly increasing in $\phi_{f,i}$ above $\bar{F}$, so that all individuals with comparative advantage and above-threshold talent in f always enter.*

Proof. Consider the utility of an individual whose highest talent is in f and with $\phi_{f,i} > \bar{F}$:

$$u_i = \phi_{f,i} + \frac{(1-\bar{F}^3)}{3} - \frac{3}{4}\left(\bar{F} + \frac{1}{1+\bar{F}+\bar{F}^2}\right) - (1-\phi_{f,i})^2$$

For these individuals $\bar{F}$ is a constant, so the derivative of u_i w.r.t. $\phi_{f,i}$ becomes:

$$\frac{\partial u_i}{\partial \phi_{f,i}} = 1 - 2(\phi_{f,i} - 1) = 3 - 2\phi_{f,i}$$

It immediately follows that this derivative is always positive for $0 \leq \phi_{f,i} \leq 1$. □

Jointly, propositions 1 and 2 show that there are no individual incentives to deviate from symmetrical talent thresholds and that these therefore constitute an equilibrium.

Uniqueness

We now show that the above solution is the unique equilibrium for the model without distortions. To do so, we illustrate that only a symmetrical interior solution can be an equilibrium. Specifically, we show that all corner solutions and potential solutions with separate thresholds for men and women generate individual incentives to deviate.

Proposition 3. *There cannot be an equilibrium in which there is entry into fewer than all f (i.e. no equilibrium exists where at least one $\bar{F} \geq 1$).*

Proof. If no one is active in f, the utility of an individual with talent $\phi_{f,i}$ considering entry into f is given by:

$$u_i = \phi_{f,i} + \phi_{f,i} - \frac{1}{2N} - (1-\phi_{f,i})^2 = 2\phi_{f,i} - (1-\phi_{f,i})^2$$

That is, the average talent of scientists in f would be equal to individual i's talent and the share of individuals active in f would be equal to one over twice the total population size, irrespective of whether i is male or female. For large N,[32] this term converges to zero and the utility of the potential first entrant becomes $u_i = 2\phi_{f,i} - (1-\phi_{f,i})^2$. The first entrant's utility is thus positive for any $\phi_{f,i} > 0.268$ and everyone above this talent level will always enter. □

Proposition 4. *There cannot be an equilibrium in which everyone enters one field (i.e. no equilibrium exists where $\bar{F} \leq 0$).*

Proof. If $\bar{F} \leq 0$, the utility of the least talented entrant ($\phi_{f,i} = 0$) in f is given by:

$$u_i = 0 + \frac{1}{2} - (1 - F(\bar{F})) - 1$$

That is, the individual's own talent is zero, the average talent of scientists in f is given by the population average of $\phi_{f,i} = \frac{1}{2}$, the share of individuals in f depends on thresholds in other fields but always falls into the $[0,1]$ interval, and lastly i's cost to enter f are equal to $(1-0)^2 = 1$. It immediately follows that this utility is always negative, i prefers the outside option, and there cannot be an equilibrium where everyone enters. □

Proposition 5. *There cannot be an equilibrium in which minimum talent thresholds differ for male and female scientists (i.e. no equilibrium exists where at least one $\bar{F}_m \neq \bar{F}_w$).*

Proof. Consider the utility of a woman i with talent $\phi_{f,i}$ under gender specific entry talent thresholds, where $0 \leq \bar{F}_m \leq \phi_{f,i} \leq \bar{F}_w \leq 1$. By this definition, the talent threshold for women prevents i from entering f while less talented men do enter f.

$$\begin{aligned} u_i = \phi_{f,i} + \theta_f \frac{(1-\bar{F}_w^3)}{3} + (1-\theta_f)\frac{(1-\bar{F}_m^3)}{3} \\ - \omega\frac{3}{4}\left(\bar{F}_w + \frac{1}{1+\bar{F}_w+\bar{F}_w^2}\right) - (1-\omega)\frac{3}{4}\left(\bar{F}_m + \frac{1}{1+\bar{F}_m+\bar{F}_m^2}\right) - (1-\phi_{f,i})^2 \end{aligned}$$

By the definition of $\bar{F}_m$, this utility is equal to zero for $\phi_{f,i} = \bar{F}_m$. It then immediately follows, that u_i is positive for all $\phi_{f,i} > \bar{F}_m$. It is therefore profitable for i to enter and asymmetric thresholds cannot be an equilibrium. □

Jointly, propositions 3 to 5 show that the interior solution with symmetric thresholds is the unique equilibrium for this model.

Model with Entry Bias

The modified system of equations is as follows:

[32] We analyse the biographies of 70,039 scientists, so that $1/2N = 0.000007$.

$$w(\bar{P}_w) - c(\bar{P}_w) - \alpha = w(\bar{N}_w) - c(\bar{N}_w) - \alpha \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{N}_m) - c(\bar{N}_m) \tag{52}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - \alpha = w(\bar{S}_w) - c(\bar{S}_w) - \alpha \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{53}$$

$$w(\bar{N}_w) - c(\bar{N}_w) - \alpha = w(\bar{S}_w) - c(\bar{S}_w) - \alpha \qquad w(\bar{N}_m) - c(\bar{N}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{54}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - \alpha = 0 \qquad w(\bar{P}_m) - c(\bar{P}_m) = 0 \tag{55}$$

$$w(\bar{N}_w) - c(\bar{N}_w) - \alpha = 0 \qquad w(\bar{N}_m) - c(\bar{N}_m) = 0 \tag{56}$$

$$w(\bar{S}_w) - c(\bar{S}_w) - \alpha = 0 \qquad w(\bar{S}_m) - c(\bar{S}_m) = 0 \tag{57}$$

$$\frac{(3 - \bar{P}_w^3 - \bar{N}_w^3 - \bar{S}_w^3)}{(6 - \bar{P}_w^3 - \bar{P}_m^3 - \bar{N}_w^3 - \bar{N}_m^3 - \bar{S}_w^3 - \bar{S}_m^3)} = 0.06 \tag{58}$$

Where equation (58) captures the share of women that we observe in MoS (1956).[33]

Model with Field Choice Bias

We set up our system of equations as follows:

$$w(\bar{P}_w) - c(\bar{P}_w) - \beta = w(\bar{N}_w) - c(\bar{N}_w) - \beta \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{N}_m) - c(\bar{N}_m) \tag{59}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - \beta = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{60}$$

$$w(\bar{N}_w) - c(\bar{N}_w) - \beta = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{N}_m) - c(\bar{N}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{61}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta = 0 \qquad w(\bar{P}_m) - c(\bar{P}_m) = 0 \tag{62}$$

$$w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta = 0 \qquad w(\bar{N}_m) - c(\bar{N}_m) = 0 \tag{63}$$

$$w(\bar{S}_w) - c(\bar{S}_w) - \alpha = 0 \qquad w(\bar{S}_m) - c(\bar{S}_m) = 0 \tag{64}$$

$$\frac{(3 - \bar{P}_w^3 - \bar{N}_w^3 - \bar{S}_w^3)}{(6 - \bar{P}_w^3 - \bar{P}_m^3 - \bar{N}_w^3 - \bar{N}_m^3 - \bar{S}_w^3 - \bar{S}_m^3)} = 0.06 \tag{65}$$

$$\frac{9(-2 + \bar{P}_w^3 + \bar{N}_w^3)}{(-2 + \bar{P}_m^3 + \bar{N}_m^3)(-3 + \bar{P}_m^3 + \bar{N}_m^3 + \bar{S}_m^3)(-3 + \bar{P}_w^3 + \bar{N}_w^3 + \bar{S}_w^3)} = 0.68 \tag{66}$$

Where equation (66) forces the model to match the relative distribution of men and women

[33] The LHS in equation (58) is a simplified expression for $\frac{(1-\bar{P}_w^3)/3+(1-\bar{N}_w^3)/3+(1-\bar{S}_w^3)/3}{(1-\bar{P}_m^3)/3+(1-\bar{H}_w^3)/3+(1-\bar{N}_m^3)/3+(1-\bar{N}_w^3)/3+(1-\bar{S}_m^3)/3+(1-\bar{S}_w^3)/3}$

between physical and social science research fields.[34] [35]

Model with Patenting Bias

We set up our system of equations as follows:

$$w(\bar{P}_w) - c(\bar{P}_w) - \gamma = w(\bar{N}_w) - c(\bar{N}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{N}_m) - c(\bar{N}_m) \tag{67}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - \beta - \gamma = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{68}$$

$$w(\bar{N}_w) - c(\bar{N}_w) - \beta = w(\bar{S}_w) - c(\bar{S}_w) \qquad w(\bar{N}_m) - c(\bar{N}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{69}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - \alpha - \beta - \gamma = 0 \qquad w(\bar{P}_m) - c(\bar{P}_m) = 0 \tag{70}$$

$$w(\bar{N}_w) - c(\bar{N}_w) - \alpha - \beta = 0 \qquad w(\bar{N}_m) - c(\bar{N}_m) = 0 \tag{71}$$

$$w(\bar{S}_w) - c(\bar{S}_w) - \alpha = 0 \qquad w(\bar{S}_m) - c(\bar{S}_m) = 0 \tag{72}$$

$$\frac{(3 - \bar{P}_w^3 - \bar{N}_w^3 - \bar{S}_w^3)}{(6 - \bar{P}_w^3 - \bar{P}_m^3 - \bar{N}_w^3 - \bar{N}_m^3 - \bar{S}_w^3 - \bar{S}_m^3)} = 0.06 \tag{73}$$

$$\frac{9(-2 + \bar{P}_w^3 + \bar{N}_w^3)}{(-2 + \bar{P}_m^3 + \bar{N}_m^3)(-3 + \bar{P}_m^3 + \bar{N}_m^3 + \bar{S}_m^3)(-3 + \bar{P}_w^3 + \bar{N}_w^3 + \bar{S}_w^3)} = 0.68 \tag{74}$$

$$\frac{(\bar{P}_m^3 + \bar{N}_m^3 - 2)(\bar{P}_w^3 - 1)}{(\bar{P}_m^3 - 1)(\bar{P}_w^3 + \bar{N}_w^3 - 2)} = 0.25 \tag{75}$$

Where equation (75) enforces the relative participation of men and women in patenting in the physical sciences.[36] [37]

[34] The relative participation is given by dividing the share of physical scientists among women by the share of physical scientists among men, $0.50/0.74 = 0.68$.

[35] The LHS in equation (66) is a simplified expression for:
$\left(\frac{(1-\bar{P}_w^3)/3+(1-\bar{N}_w^3)/3}{(1-\bar{P}_w^3)/3+(1-\bar{N}_w^3)/3+(1-\bar{S}_w^3)/3}\right) \times \left(\frac{(1-\bar{P}_m^3)/3+(1-\bar{N}_m^3)/3}{(1-\bar{P}_m^3)/3+(1-\bar{N}_m^3)/3+(1-\bar{S}_m^3)/3}\right)^{-1}$

[36] The LHS in equation (75) is a simplified expression for:
$\left(\frac{(1-\bar{P}_w^3)/3}{(1-\bar{P}_w^3)/3+(1-\bar{N}_w^3)/3}\right) \times \left(\frac{(1-\bar{P}_m^3)/3}{(1-\bar{P}_m^3)/3+(1-\bar{N}_m^3)/3}\right)^{-1}$

[37] $0.09/0.38 = 0.25$