The Ms. Allocation of Talent

Moritz Lubczyk, Rockwool Foundation Berlin, and
Petra Moser, NYU, NBER, and CEPR*

December 10, 2024

Despite improvements in the allocation of talent, women continue to be underrepresented in innovation. This paper investigates whether changes in the allocation of talent could close this innovation gender gap. Linking 70,000 scientists with their patents, we find that the underrepresentation of women 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 persistent gender gap in innovation. We interpret these findings through the lens of a Roy (1951) model of field choice with gender distortions. To test the model's predictions and to identify the causal effects of gender differences in the allocation of talent, we exploit an exogenous shock in female representation due to WWII. As men enlisted in the war, 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. Using data on PhDs and elite education, we show that female scientists are positively selected and that selection decreased during WWII when more women entered science. Counterfactual estimates imply that if women were as likely to work in STEM fields as men, the innovation gender gap would close in 38, rather than 118 years.

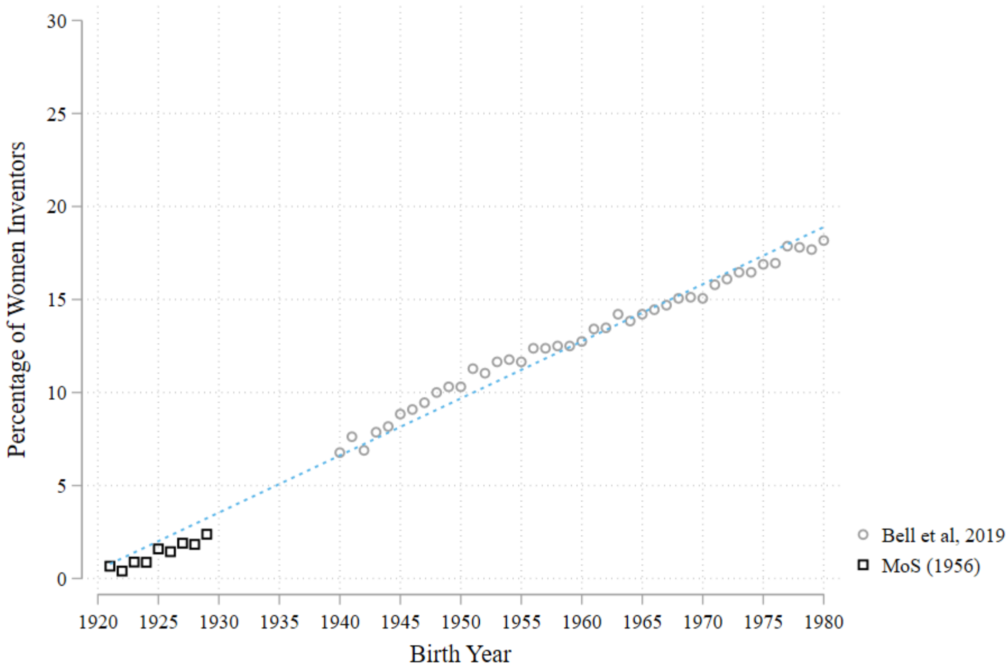
Keywords: Innovation, Science, Gender, and Misallocation

*We thank Joseph Altonji, Nava Ashraf, David Card, Christian Dustmann, Martin Fiszbein, Marita Freimane, Jacob French, Jacopo Gambato, Jessica Pan, Fabian Waldinger, and seminar participants at Barnard, Bayreuth, LSE, the DAE and Gender Economics Summer Institutes, Northwestern, Nottingham, NYU, Católica, Rockwool Foundation Berlin, Rockefeller University, WZB Berlin, ZEW Mannheim, and Zurich for helpful comments and discussions.

As talented women and Black men have entered high-skilled professions, improvements in the allocation of talent have raised productivity (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 a comprehensive compendium of scientists, the *American Men of Science* (MoS 1956). To analyze changes in the innovation gender gap, we 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).

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



Note: Female inventor share combining data from MoS (1956) and Bell et al. (2019). We define the female 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 during their career.

To investigate the drivers of this persistent gender gap, we estimate OLS regressions of the probability that a scientist becomes an inventor on an indicator for female scientists and control variables for productivity, industry employment, educational attainment, family status, and research fields. In these regressions, a scientist's research focus explains most of the gender differences in patenting. A Kitagawa-Oaxaca-Blinder decomposition of the innovation gender gap confirms these results. 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 higher in the social sciences (9%), but patenting in the social sciences is rare.

We present a Roy (1951) model of field choice with gender distortions to interpret these results. In the model, men and women decide whether to become scientists and which field to enter based on their comparative advantage. We assume that men and women draw from the same distribution of talent but that women face distortions that increase their cost of specializing in fields with patenting. Taking the model to the data, we show that the observed innovation gender gap implies that women face much higher talent thresholds to enter research fields with patenting than men.

The Roy model yields four key predictions. First, due to distortions, women are underrepresented in fields with patenting. Second, women in fields with patenting are positively selected relative to men. Third, as men are pulled from the labor force, more women enter science. Fourth, as more women enter science, these additional female entrants continue to be positively selected relative to men, but to a lesser degree.

Data on educational attainment and research productivity confirm these predictions. Female scientists are significantly more likely to have earned a PhD, with an 80% share of PhDs compared with a 70% among men. Women are also more likely to earn degrees from elite, Ivy Plus institutions, with a 44% share compared with a 35% share for men. Moreover, female scientists produce better patents: Patents by female scientists are 3.0 percentage points more likely to be among the top 5 percent most highly cited patents in their discipline and 4.4 percentage points more likely to be among the top 10 percent than patents by male scientists. Selection was weaker during WWII, when the shortage of male labor drew more women into science.

To investigate the causal effects of changes in the allocation of female talent, we exploit an exogenous shift due to WWII, when the scarcity of male scientists pulled female scientists into male-dominated, patent-intensive fields. By linking the MoS with the *U.S. World War II Army Enlistment Records, 1938-1946*, we are able to measure variation in 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 peaked with a 92% increase in 1942 and remained high until 1950. Fields with high (above the median) enlistment shares experienced a 40% increase in female entry.

Using field-level variation in exposure to enlistment as an instrument for female entry, we find that, for every five women entering the physical sciences, one additional woman becomes an inventor. In a complementary set of individual-level instrumental variable (IV) regressions, 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 percentage points.

Incorporating IV estimates in reweighting counterfactuals (DiNardo, Fortin, and Lemieux 1996), we explore whether a more equal allocation of talent would help close the innovation gender gap. Counterfactuals 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. If women entered STEM at the same rate as men, gender parity in innovation would be reached in 38 years rather than the 118 years implied by current rates of convergence (Bell et al. 2019).

These findings contribute causal evidence on the economic consequences of the misallocation of talent to the literature on inequality in innovation. Hsieh et al. (2019) show that 20 to 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 high-skilled occupations. Focusing on gender inequality among university scientists, however, Iaria, Schwarz, and Waldinger (2022), document persistent gender gaps in hiring, publishing, and promotions, even though the share of women at elite universities has increased from 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 estimating the causal effects of improvements in the allocation of talent and by quantifying the distortions that lead to underrepresentation.

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

[1] Acemoglu et al. (2004) show that states with a 10% higher mobilization rate experienced a 9.8% increase

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

Until WWII, female scientists were relegated primarily to "women's work," even though they had succeeded in gaining access to nearly all universities in the United States and Germany. A separate labor market for women emerged in the 1880s and 1890s, when women began to seek scientific employment in significant numbers, and was firmly established before WWI. According to Rossiter (1982, p.51), "this practice of 'sex segregation' was usually justified with the essentially conservative rhetoric that women had 'special skills' or 'unique talents' for certain fields or kinds of work."

With sex segregation in place, countless women failed to find employment as scientists; 138 of these women are listed in the MoS (1938) as "unemployed." Other women chose to work in more traditional women's occupations such as "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, p.262f). For instance, Gertrude B. Elion, a co-recipient of the 1988 Nobel prize in medicine, worked as a high school teacher in New York until 1942.

Limits on the participation of women in science became impractical after September 1940, when all men between the age of 21 and 35 became subject to the WWII draft lottery under the Selective Training and Service Act of 1940 (Pub. L. 76-783, 54 Stat. 885). Amendments in 1941 and 1942 extended the draft to cover all men between the age of 18-45 (Goldin and Olivetti 2013; Acemoglu et al. 2004).

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.

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 in 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.

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).[2] Despite its name, the MoS includes male and female scientists in Canada and the United States. A comprehensive database of scientists, this dataset was 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. In the MoS, Catell created "for the first time a fairly complete survey of the scientific activity of a country in a given period" (Cattell 1906, p.v). 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 are 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 and colleges, industrial laboratories and other institutions" (Cattell 1956, p. iv). 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."

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 benchmarking exercise that uses graduate 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.

Data on marriages and children allow us to control for variation in the family status of male and female scientists. 81.7% of scientists are married, and 71.1% have at least one child.

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 scientist's place of birth allow us to control for the influence of a person's childhood environment. 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 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 were born in other

[2]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.

countries (10.1%). We use these data to control for birth places at the level of US states and a dummy for foreign-born scientists.

Detailed information on career histories allows us to distinguish academic scientists from industry scientists. We use job titles, such as "professor," "research fellow," or "instructor," to identify 52,785 academic scientists, who worked in academia at least once. Our data include 3,525 female academic scientists (87.7% of female scientists) and 49,260 male academic scientists (74.6% of male scientists). 17,254 scientists work exclusively in industry; 495 of these scientists were women and 16,759 were men.

We use the year when a scientist started their first job to identify the year when they entered science. For example, Yolanda Tota Pratt started her first job as a lecturer at Barnard College in 1941, the year before she obtained her PhD at Columbia.

> **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 implement a *k*-means matching algorithm (from Kim and Moser (2024), with *k*=100) to the text that describes each scientist's research topics and discipline. Research topics are known for 96.4% of 82,094 scientists in the MoS (1956); disciplines are known 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 extend these field assignments by creating a nested structure that assigns each of the 100 research fields to one of the 3 sciences. Specifically, we assign 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. We refer to research fields in the physical and biological sciences as STEM fields.

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. This approach exploits an improved matching process from Moser, Parsa, and San (2024) 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.[3] For patents in the physical sciences, we further obtain patent forward citations as a measure of patent quality.[4] We focus on single-inventor patents and define highly cited patents as patents among the top 5%, 10%, and 25% most-cited patents in their discipline by publication year. 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 names), 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 names upon marriage. We are, however, unlikely to miss many patents: Just 38.8% of female scientists married (1,560 women), and just 466 married women worked in the physical sciences, where scientists are most likely to patent. 91.0% of married female scientists (1,418 women) report their middle name, which, for married women, is typically the maiden name. In the physical sciences, 93.1% of married female scientists (434 women) report their maiden name – only 32 married women do not. We use this information to match women with the patent record. For example, we link Julianne Prager (née Heller) to seven patents:

[3]Moser et al. (2024) use these data to investigate the effects of the 1920s quota act on innovation.
[4]Nguyen and Moser (2024) construct these data to study the effects of Nazi scientists on US innovation.

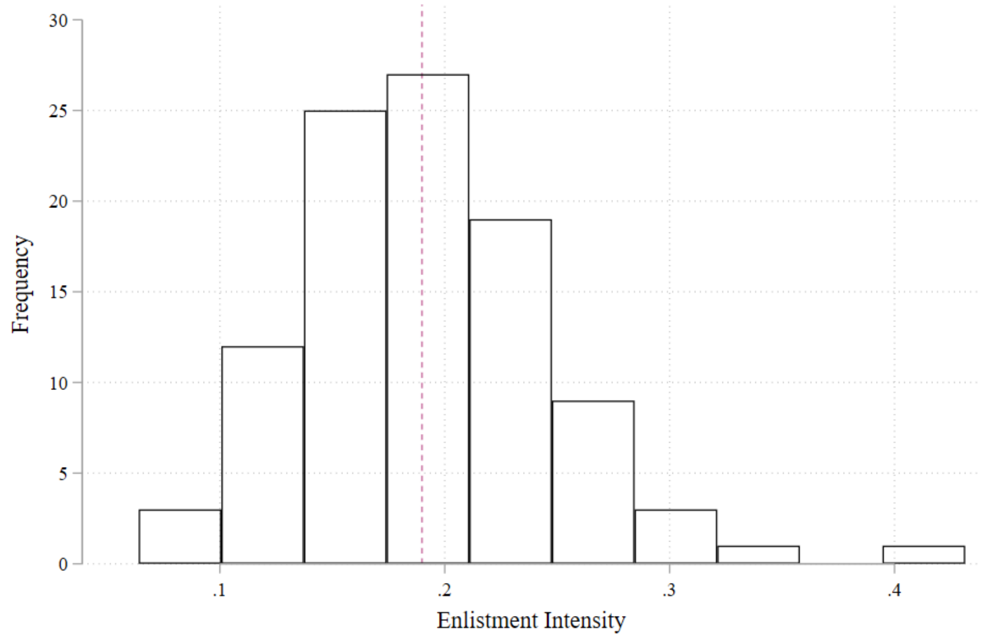
PRAGER, DR. JULIANNE H(ELLER), (MRS. STEPHEN PRAGER), 1230 Rose Vista Court, St. Paul 13, Minn. ORGANIC CHEMISTRY. Boston, Mass, June 5, 24; m. 48. Sc.B, Brown, 46; Ph.D.(org. chem), Cornell, 53. Research assoc. biochem, Utah, 51-52; SR. CHEMIST, MINN. MINING & MFG. CO, 52- Chem. Soc. Synthesis of macrocyclic ketones, C^{14}-labelled steroids and polymers.

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

Matching Scientists with WWII Enlistment Records

To measure variation in exposure to enlistments across fields, we exploit a record linkage from Chen and Moser (2024), who link scientists with 9,025,234 men and women enlisted in the US Army and Air Force between 1938-46.[5] 10,285 scientists enlisted in WWII; 19% of 54,424 draft-eligible men between the ages of 18 and 44 between 1938 and 1946. Counts of enlisted scientists increased during the war and declined in 1945, 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 (Figure B1) and 19% of male scientists in the median field enlisted (Figure 2).

Figure 2: Enlistment across Fields



Note: Share of enlisted draft-eligible scientists across research fields. Scientists matched with enlistment records from Chen and Moser (2024). Draft-eligible scientists defined as men born 1897-1927.

[5]Chen and Moser (2024) use these data to investigate the long-run effects of exposure to combat on the research productivity of scientists.

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).[6] We match scientists in the MoS to their MAG author identifier using their 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.

On average, scientists produced 16.6 publications between 1900 and 1970. With 980 journal 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. The Innovation Gender Gap

Why are women underrepresented in innovation? To investigate this question, we start by investigating alternative channels for the gender gaps in patenting for a population of highly skilled scientists who have the expertise required to patent.

OLS Estimates 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. The matrix X_i includes potential explanatory variables for the innovation gender gap. To evaluate the explanatory power of alternative channels, we observe the change in β_1 conditional on subsets of explanatory variables in X_i and the interaction between X_i and $\mathbb{1}[Female]_i$. Table A.1 provides summary statistics across all traits.

At baseline, female scientists are 19.4 pp. less likely to patent than male scientists born in the same year and state (Table 1, column 1). Birth year and state fixed effects control for variation in patenting across birth cohorts and states, e.g., if women who grew up in locations with more female inventors are more likely to patent (Bell et al. 2019).

The control variable *PhD* is an indicator for scientists who hold a doctoral degree, to measure differences in qualifications. Controlling for PhD does not change the gender innovation gap (19.5 pp., Table 1, column 2). Women are more likely to have a PhD (79.8%)

[6] Moser and Parsa (2024) create this matching to investigate the effects of McCarthyism on American science. MAG was updated weekly until December 2021; we use the version from August 20, 2020.

than men (70.1%), but a PhD does not affect the probability of patenting for men or women.

Next, we investigate whether female scientists patent less because they are less productive. Specifically, we use *publications* as a measure of productivity that is unrelated to patent-specific gender barriers.[7] We measure publications as the natural logarithm of a scientist's publications.[8] OLS estimates, however, show that 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 large, at 16.3 pp., with controls for publications (Table 1, column 3). The additional increase in patenting for scientists with more publications is 1.2 pp. larger for men.

The variable *industry* distinguishes scientists working in industry (rather than in academia, where promotions are typically tied to publications instead of patents). Controlling for industry employment reduces the gender gap to 15.5 pp. (Table 1, column 4). Scientists of either gender are more likely to patent if they work in industry, but women are 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 pp.) than for women (+6.1 pp., Table 1, column 4).

We also control for variation in patenting as a result of marriage and children. For example, *married* women may patent less if joint location decisions prioritize the husband's career ("the tied migration of working wives," Mincer 1978) or as a result of marriage bars (Goldin 1988).[9] In addition to marriage, children may reduce the productivity of mothers in science (Kim and Moser 2024). Controlling for marital status reduces the gender gap to 15.0 pp. (Table 1, column 5); controlling for children leaves it at 15.5 pp. (Table 1, column 6). Married scientists and those with children are more likely to patent if they are male but less likely if they are female. Male scientists are 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%).

Finally, we compare patenting rates for scientists in the *physical* and *biological sciences* with rates for social scientists as the omitted category. OLS estimates indicate that gender differences in the allocation of talent across the physical, biological, and social sciences explain, by far, the largest share of the innovation gender gap (Figure 3). Controlling

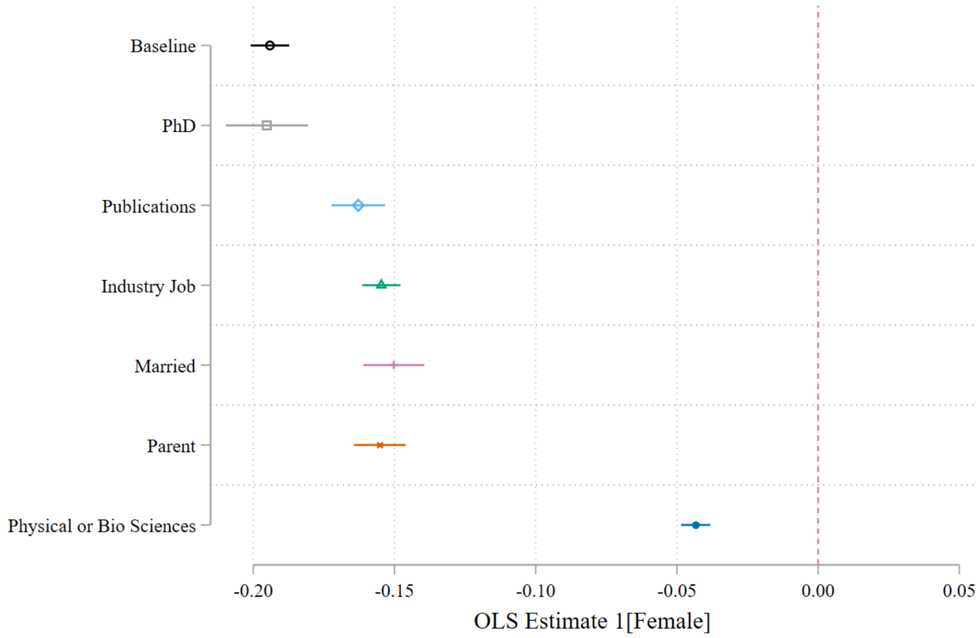
[7]In addition to direct discrimination in the patent system (Jensen, Kovács, and Sorenson 2018; Pairolero, Toole, DeGrazia, Teodorescu, and Pappas 2022), women may also patent less because they are more likely to abandon patent applications (Aneja, Reshef, and Subramani 2024), or because they get less credit for their work in group settings (Sarsons, Gërxhani, Reuben, and Schram 2021) and are not named on the patent document.

[8]In the main specification, we add 1 to the count of publications. Table A.3 shows that our results are robust to measuring publications as raw counts or using the inverse hyperbolic sine transformation instead.

[9]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. 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 possibility."

for the working in the physical and biological sciences reduces the innovation gender gap to 4.3 pp. (Table 1, column 7).

Figure 3: OLS Prediction of the Innovation Gender Gap



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

Table 1: OLS Decomposition of the Gender Innovation Gap

	(1) Baseline	(2) PhD	(3) Publications	(4) Industry Job	(5) Married	(6) Parent	(7) Physical or Bio Sciences
Female	-0.194***	-0.195***	-0.163***	-0.155***	-0.150***	-0.155***	-0.043***
	(0.003)	(0.007)	(0.005)	(0.003)	(0.005)	(0.005)	(0.003)
PhD		-0.005					
		(0.004)					
Female × PhD		0.002					
		(0.008)					
Publications			0.023***				
			(0.001)				
Female × Publications			-0.012***				
			(0.003)				
Industry				0.193***			
				(0.004)			
Female × Industry				-0.132***			
				(0.014)			
Married					0.049***		
					(0.004)		
Female × Married					-0.057***		
					(0.007)		
Children						0.048***	
						(0.004)	
Female × Children						-0.063***	
						(0.008)	
Physical Science							0.334***
							(0.003)
Female × Phys. Sc.							-0.237***
							(0.009)
Biological Science							0.057***
							(0.003)
Female × Bio. Sc.							-0.035***
							(0.005)
Constant	0.232***	0.236***	0.183***	0.184***	0.191***	0.196***	0.042***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)
Birth year FE	✓	✓	✓	✓	✓	✓	✓
Birth state FE	✓	✓	✓	✓	✓	✓	✓
Mean Y	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
R-squared	0.022	0.022	0.026	0.060	0.023	0.024	0.145

Note: Robust errors in parentheses. In the main specification, we measure publications as the natural logarithm plus one. We show that this is robust to alternative measures (Table A.3). *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Kitagawa-Oaxaca-Blinder Decomposition of the Innovation Gender Gap

For an alternative decomposition of the innovation gender gap, we separate the effects of differential predictor endowments (here, the levels of scientists' traits) and differential predictor returns (here, the coefficients associated with scientists' traits) in the spirit of Kitagawa (1955), Oaxaca (1973), and Blinder (1973). To conduct this decomposition, we estimate separate OLS regressions for male and female scientists:

$$\mathbb{1}[Inventor]_{i,g} = \sum \beta_{j,g} X_{j,g} + \varepsilon_{i,g} \tag{2}$$

where the indicator variable $\mathbb{1}[Inventor]$ equals 1 if scientist i is an inventor, g indexes gender, and X_j are variables that influence whether scientist i is an inventor: indicators for a PhD, industry employment, marriage, parenthood, and employment in the physical or biological sciences, as well as the logarithm of publications, birth years, and birth states.

We use these OLS estimates to decompose the innovation gender gap:

$$\overline{\text{Inventor}}_m - \overline{\text{Inventor}}_f = \sum \beta_{j,m} \times (\bar{X}_{j,m} - \bar{X}_{j,f}) + \sum \bar{X}_{j,f} \times (\beta_{j,m} - \beta_{j,f}) \tag{3}$$

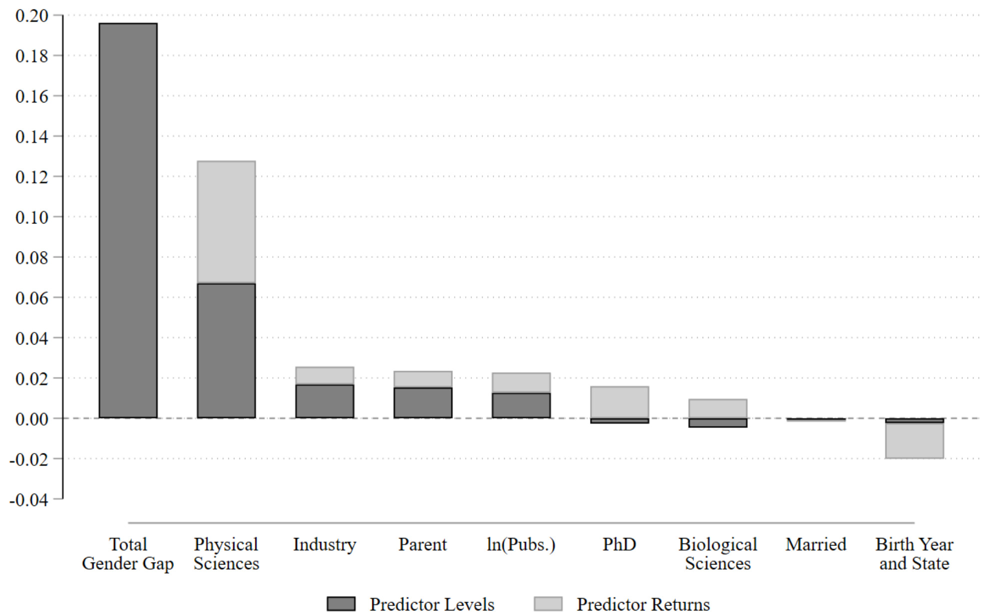
The first term on the right hand side, $\sum \beta_{j,m} \times (\bar{X}_{j,m} - \bar{X}_{j,f})$, captures the share of the innovation gender gap that is attributable to gender differences in levels of predictors. The second term, $\sum \bar{X}_{j,f} \times (\beta_{j,m} - \beta_{j,f})$, captures gender differences in the returns to predictors. Assuming that men do not face discrimination, we set the coefficients for male scientists as the reference coefficients for female scientists.[10]

This decomposition shows that differences in the allocation of male and female scientists across fields account for the largest share of the innovation gender gap (Figure 4). The unconditional innovation gender gap is 19.6 percentage points; of this gap, 12.8 percentage points (65.1%) are attributable to employment in the physical sciences. Specifically, 6.7 percentage points are due to women being less likely to work in the physical sciences, and 6.1 percentage points are due to women experiencing lower returns to entering the physical sciences. All other predictors account for smaller shares of variation: 1.3 percentage points are due to having a PhD, and in particular to women earning lower returns on PhD degrees. Publications account for 2.3 percentage points, industry employment for 2.6 percentage points, and motherhood for 2.4 percentage points.[11]

[10] Oaxaca (1973) describes the choice of assumption about reference coefficients as the 'index number problem.' While we assume that men's returns are unaffected by discrimination for simplicity here, we provide evidence that men in fact benefit from distortions keeping women out of science when we develop our Roy-model of occupational choice in science. See Cotton (1988) for a decomposition that considers positive effects on the reference group.

[11] Birth states and years contribute negatively to the innovation gender gap (-2.0 percentage points), which suggests that the gap would be even larger if female scientists were distributed similarly to men across birth states and years and that women do not benefit from place and cohort effects to the same extent as men.

Figure 4: Kitagawa-Oaxaca-Blinder Decomposition of the Innovation Gender Gap



Note: Decomposition of innovation gender gap in predictor levels and predictor returns. Predictors sorted by attributable share of gender gap. The last column combines birth year and state fixed effects.

4. Roy (1951) Model of Field Choice

We present a simple Roy (1951) model to investigate the effects of field choice distortions on the innovation gender gap. Scientists do research in two types of fields. In fields P (mechanical engineering or other fields in the physical sciences), patenting is common, while in fields S (anthropology or another social science), patenting is rare. Individuals i choose between pursuing science (in either field) and working in another job based on their comparative advantage. We normalize the utility of working outside of science to zero and assume that men and women face the same value of outside options. A share ω of the population are women; $\omega = 0.5$ in the baseline model.

Talent, Wages, and Costs

Let $\phi_{f,i}$ denote individual i's talent in field f, which can be either a field with patenting (P) or without patenting (S). Each individual is endowed with a vector $\Phi_i = \{\phi_{S,i}, \phi_{P,i}\}$ of i.i.d standard uniform talents for research in patenting ($\phi_{P,i}$) and non-patenting ($\phi_{S,i}$) fields. The distribution of talent is identical for women and men. An individual's talent determines both the wage they earn from specializing in f and their cost of specialization c.

After individuals see their talent draws across all fields, they choose between entering science and the outside option. Individuals enter science in a specific field that they choose when they enter. Individuals maximize their utility:

$$\max_f \; U(f) = \max[w(\phi_{f,i}) - c(\phi_{f,i}),\, 0] \tag{4}$$

where $w(\phi_{f,i})$ is the wage function and c captures the cost of individual i specializing in field f. The cost of specialization $c(\phi_{f,i})$ is convex in an individual's talent so that more talented scientists face lower costs of entry: $c(\phi_{f,i}) = (1-\phi_{f,i})^2$. Wages are endogenous. They increase in individual i's own talent in f, so that more talented scientists earn higher payoffs. Wages decrease as a larger share of the population are scientists in field f and the field becomes more crowded.

$$w(\phi_{f,i}) = \phi_{f,i} - \omega\frac{(1-\bar{f}_w^2)}{2} - (1-\omega)\frac{(1-\bar{f}_m^2)}{2} \tag{5}$$

$\omega\frac{(1-\bar{f}_w^2)}{2}$ and $(1-\omega)\frac{(1-\bar{f}_m^2)}{2}$ denote the share of women and men, respectively, who become scientists in f. The talent threshold $\bar{f}$ is the minimum talent required to enter f.[12] The share of entrants in field f 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 threshold for men.

Equilibrium

We solve the model for the equilibrium talent thresholds at which the marginal woman or man is indifferent about entering a field f. We use the utility function to define a system of simultaneous equations. Equations (6) to (8) describe women's indifference conditions. At the talent thresholds $\bar{P}_w$ and $\bar{S}_w$ the marginal woman is indifferent between entering P or S and not entering science (and take the outside option with a payoff of zero).

In fields with patenting, payoffs for female scientists are distorted by a parameter d. For instance, d may be high if women 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; Wasserman 2023), scheduling flexibility (Bronson 2015), teacher discrimination, role models (Carrell, Page, and West 2010) and work environments (Wiswall and Zafar 2018).[13] This distortion generates the innovation gender gap r; we infer d by observing r.

[12] Consider the bivariate cumulative density function F of Φ_i at $\bar{f}$. For i.i.d. standard uniform talents it is easy to see that for field f relative to field h, $1-F(\bar{f}) = P(\{\phi_f > \phi_h\} \cap \{\phi_f > \bar{f}\} = \int_{\bar{f}}^1 \int_0^{\phi_f} 1 \, d\phi_h \, d\phi_f = \frac{1-\bar{f}^2}{2}$.

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

$$w(\bar{P}_w) - c(\bar{P}_w) - d = w(\bar{S}_w) - c(\bar{S}_w) \tag{6}$$

$$w(\bar{P}_w) - c(\bar{P}_w) - d = 0 \tag{7}$$

$$w(\bar{S}_w) - c(\bar{S}_w) = 0 \tag{8}$$

Equations (9) to (11) describe the equivalent indifference conditions for men, without distortions.

$$w(\bar{P}_m) - c(\bar{P}_m) = w(\bar{S}_m) - c(\bar{S}_m) \tag{9}$$

$$w(\bar{P}_m) - c(\bar{P}_m) = 0 \tag{10}$$

$$w(\bar{S}_m) - c(\bar{S}_m) = 0 \tag{11}$$

Equation (12) states that the share of women who enter patenting fields relative to the share of men is equal to a ratio r and closes the model.

$$\left(\frac{1-\bar{P}_w^2}{2}\right)\left(\frac{1-\bar{P}_m^2}{2}\right)^{-1} = r \tag{12}$$

Solving the model analytically (Appendix C) yields the parameters:

$$\bar{S}_m = 3 - \sqrt{6} \qquad \bar{S}_w = 3 - \sqrt{6}$$

$$\bar{P}_m = \frac{\gamma - 6}{r - 3} \qquad \bar{P}_w = \theta$$

$$d = 3\theta - \frac{6 + \gamma(3 - 9r) + 22r + 8r^2}{(r-3)^2}$$

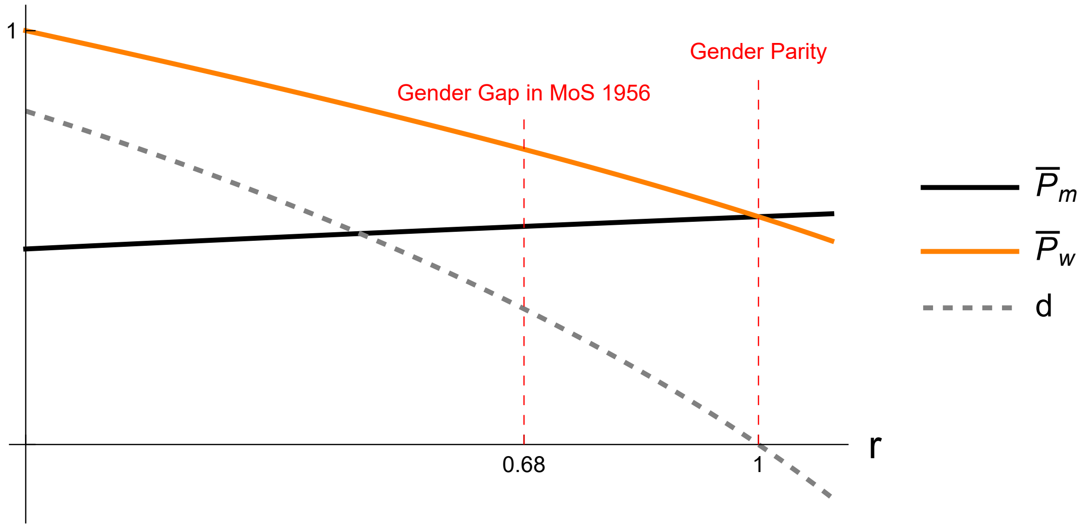
where

$$\theta = \frac{\sqrt{3}\sqrt{-1 + 3r + \frac{24r}{-3+r} - \frac{4r\sqrt{21+2r+r^2}}{-3+r}}}{\sqrt{-3+r}} \qquad \gamma = \sqrt{21 + 2r + r^2}$$

This expression allows us to infer talent thresholds $\bar{P}_w$ and $\bar{P}_m$, as well as the distortion parameter d as a function of the observed innovation gender gap r. Equilibrium talent thresholds in the non-patenting fields do not depend on r and are symmetrical for men and women.[14] In contrast, talent thresholds in the fields with patenting depend on the magnitude of the innovation gender gap for both women and men. Figure 5 illustrates this relationship:

[14] For simplicity, we assume that individuals with a comparative advantage in P and with talents below the equilibrium talent threshold take the outside option. If some of them are talented enough to enter S, their choice raises the talent threshold in S symmetrically for men and women by increasing the number of scientists.

Figure 5: Talent Thresholds and Distortion Parameter as Function of the Gender Gap



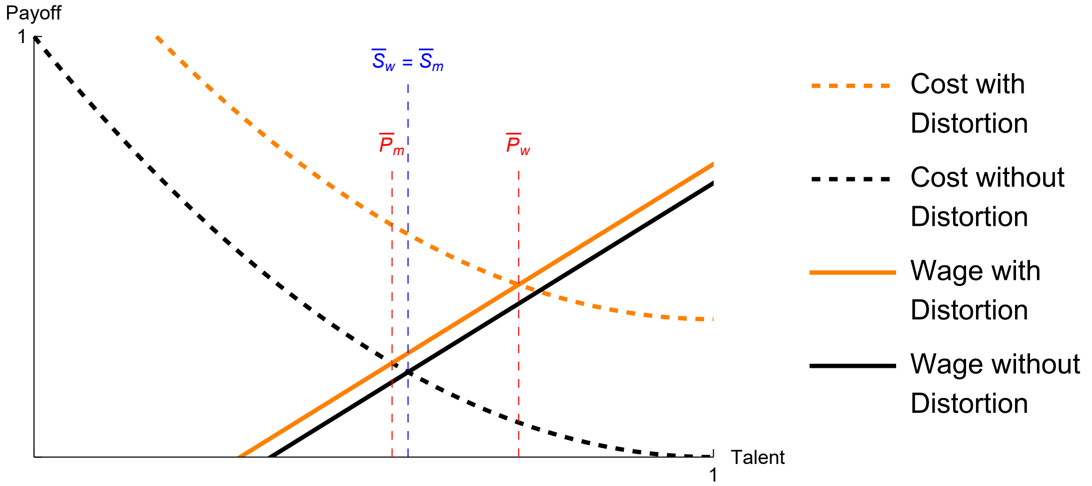
Note: $\bar{P}_m$ and $\bar{P}_w$ are the talent thresholds at which men and women are indifferent about entering fields with patenting, P. r is the innovation gender gap, measures as the ratio of the shares of women and men who enter P. At $r = 1$, women and men are equally likely to enter P and $r = 0.68$ creates the gender gap observed in the MoS (1956).

Gender parity represents the symmetrical equilibrium. At parity, the ratio of the shares of women and men who enter the patenting fields is $r = 1$. Male and female talent thresholds intersect and the implied gender distortion in the patenting fields is $d = 0$.

Deviations from gender parity result in an asymmetric equilibrium where men and women face different talent thresholds for entering P. As the innovation gap rises and the ratio of proportions of women and men who enter the patenting fields shrinks, female talent thresholds increase, male talent thresholds decrease, and the implied gender distortion rises. At the observed innovation gender gap in the MoS (1956), $r = 0.68$, women face significantly higher talent thresholds and distortions than men.[15]

[15]As Figure 5 implies, the innovation gender gap would reverse if the gender distortion were negative and women received a bonus to entering P. In this case, women's talent thresholds would be lower than men's. Our analytical solutions, however, suggest that the one-sided distortions of our model are not defined for higher values of r – high levels of underrepresentation can only be achieved by imposing additional cost on the underrepresented group and cannot be generated by assigning a bonus to the overrepresented group alone.

Figure 6: Model Equilibrium based on Observed Gender Gap



Note: $\bar{P}_m$ and $\bar{P}_w$ are the talent thresholds at which men and women are indifferent about entering fields with patenting, P. $\bar{S}_m$ and $\bar{S}_w$ are the talent thresholds for fields without patenting, S.

Gender distortions affect the equilibrium in two ways (Figure 6). First, women face higher costs when entering research fields with patenting. This increase in cost decreases the payoffs women receive at every level of the talent distribution. Second, by reducing competition, distortions raise wages in fields with patenting for both male and female scientists.

Jointly, these two effects increase the talent thresholds for women while lowering the threshold for men. As women's cost curve shifts up, the new cost and wage curves cross at a higher talent level. For men, however, the cost curve remains unchanged. Due to reduced competition, as fewer women enter, men receive higher wages at every talent level. As a result, the old cost curve and the new wage curve intersect at a lower level of talent relative to the equilibrium without distortions.

Predictions

Our model yields four predictions to guide the empirical analyses:

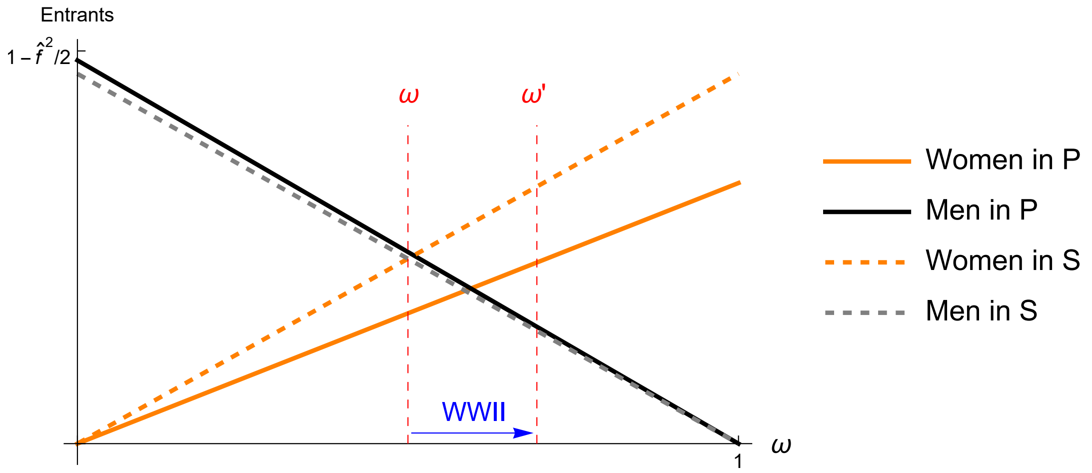
First, due to distortions, women are underrepresented in fields with patenting. Since women face higher thresholds to enter fields with patenting than men, there are fewer women above the threshold.

Second, women in science are positively selected relative to men. Because women face higher talent thresholds than men in some fields within science, the average talent of the women who enter science exceeds the average talent of men who enter science.

Third, as men are pulled from the labor force, more women enter. As men enlist in the war, the population of potential entrants changes, with a reduction in the share of of men. Equation (5) captures this effect as an increase in ω, the population share of women, which results in more women entering science. Figure 7 illustrates this shift from the initial $\omega = 0.5$

to ω'. At ω' fewer men enter and more women enter fields with and without patenting.[16]

Figure 7: Effect of WWII on Entry



Note: ω is the share of women in the population. The shift from ω to ω' is proportional to the enlistment of male scientists. Y-axis shows the population share of men and women entering fields with patenting, P, and without patenting, S, for a given value of ω under fixed distortions.

Fourth, as more women enter science, female entrants become less positively selected. While more women enter fields with patenting when men in the war, female entry cannot fully compensate for male enlistments because women continue to face field choice distortions. Graphically, the slope of women's entry curve into fields with patenting P is less steep than the slope of the male entry curve and entry curves for fields without patenting S (without distortions and with identical slopes for men and women). Since the increase in female entry into P is less than proportionate to male enlistment, the total number of entrants falls during the war. Consequently, scientists face less competition in the patenting fields and their wage increases, lowering talent thresholds for both men and women. While women continue to be positively selected relative to men, women entering during the war are less positively selected because some female entrants have talent draws below the prewar threshold when $\omega = \omega$'.

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

To investigate the causal effects of gender differences in the allocation of talent we first check whether the shortages of male scientists during WWII pulled women into science. Specifically, we estimate OLS event studies comparing changes in female entry after 1939 across fields that were differentially exposed to male enlistments.

$$Y_{kt} = \alpha_k + \phi_t + \sum_{t,t \neq 1939}^{T} \big(\mathbb{1}[T = t] \times D_k\big)\beta_t + \varepsilon_{kt} \tag{13}$$

[16] We normalize the population to 1, so that there is no effect of absolute changes in the size of the population of potential entrants.

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 long and inflexible work schedules required for laboratory work. 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 the passage of the the Selective Training and Service Act (Pub. L. 76–783, 54 Stat. 885) in October 1940. Under the assumption that, had WWII not drawn 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.

A potential challenge to the identifying assumption is that enlistment may be higher 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 female scientists before 1940 (Appendix Figure B2). OLS estimates show no significant correlation between a field's exposure to enlistment after 1940 and the number of women entering the field (Appendix Table A.2, column 1, p-value = 0.45), the number of female inventors (column 2, p-value = 0.20), or the size of the field before 1940 (column 3, p-value = 0.13).

Further investigating the identifying assumption, we check whether scientists in certain fields were more likely to enlist. Specifically, we estimate OLS regressions with enlistment as an outcome variable for men in birth cohorts between 1910 and 1927 who were most likely to enlist. We find that – conditional on marital status, parenthood, birth year, birth state, and birth year-by-birth state fixed effects – the estimated field fixed effects are consistent with random enlistment. Of 100 estimated field fixed effects, just one field (in clinical psychology) is significant at 1% and 6 are significant at 5% (Appendix Figure B3). Fields with above and below median enlistment rates also show similar trends in female entry before the draft (Appendix Figure B4). Beginning in 1940, 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 for the remainder of the sample.

Moreover, 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.

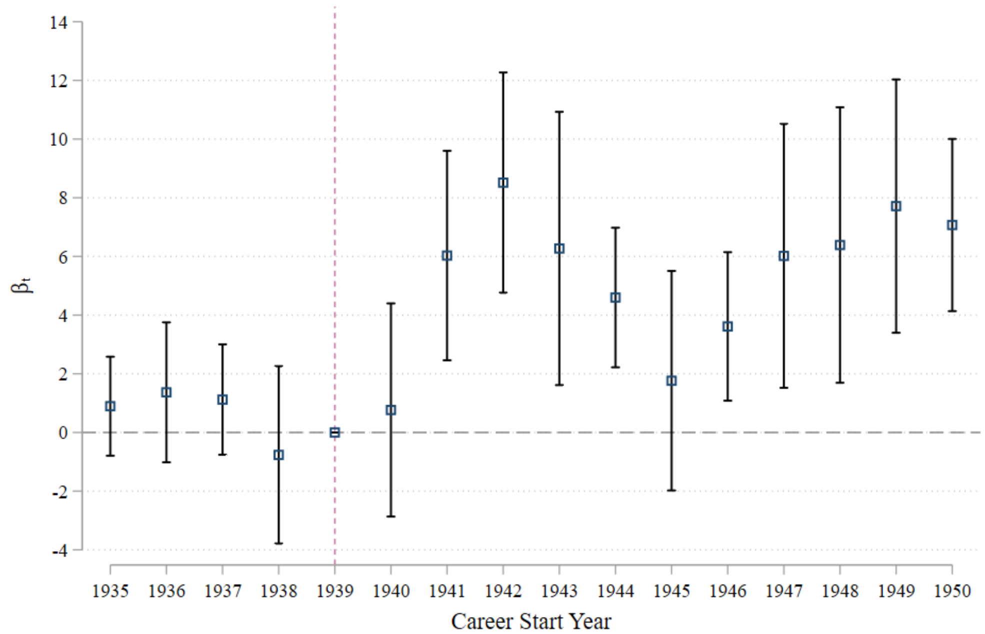
To identify the average treatment effect of enlistment on women entering treated fields we estimate difference-in-differences OLS regressions:

$$Y_{kt} = \alpha_k + \phi_t + \left(\mathbb{1}[t > 1939] \times D_k\right)\gamma + \varepsilon_{kt} \tag{14}$$

where $\mathbb{1}[t > 1939]$ is an indicator for years after the WWII draft, 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 8). 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). Even 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.

Figure 8: Event Study: 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 research fields.

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 2, 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 2, 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 regression imply a 35.7% increase in female entry for each additional 10% share in enlisted men (column 4). Estimates with the IHS indicate a 44.1% increase.

Table 2: 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	PML	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

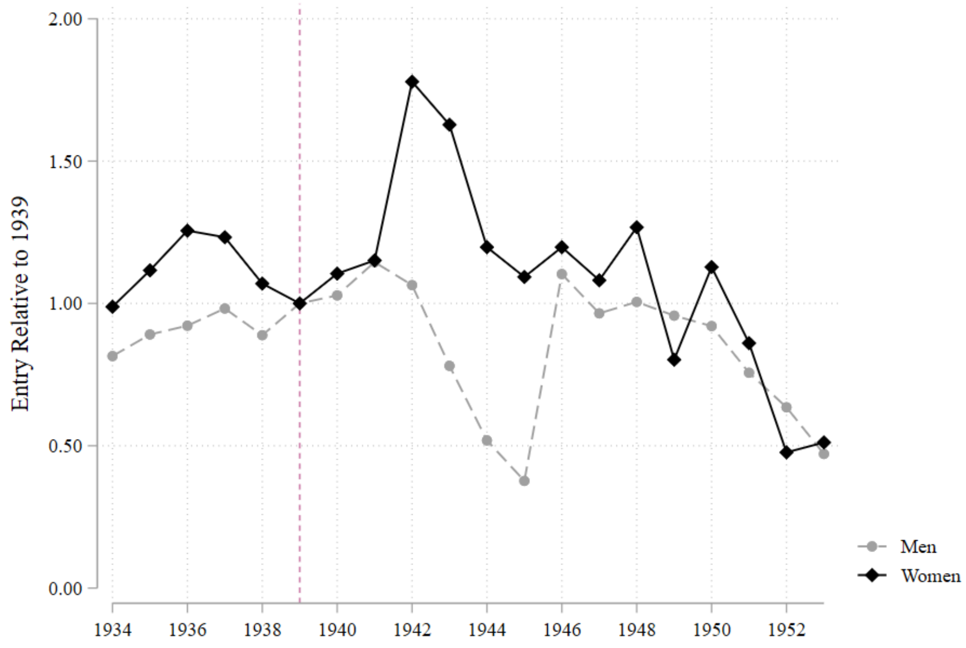
Note: OLS and Poisson estimates of two-way fixed effects difference-in-differences specifications. Treatment defined as $\mathbb{1}[Enlistment > Median]$ in Column (1) and as the share of enlisted male scientists in Columns (2), (4), and (5). Treatment defined along quartiles of the enlistment intensity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

Who are the women who became scientists during the war? We find that WWII drew both younger (below the age of 20) and older women (between the age of 28 and 40) into science (Appendix Figure B5). 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 B6). Data on career titles from the MoS indicate that many of these women came from jobs as school teachers, translators, and librarians (Appendix Figure B7).

6. Field-Level Instrumental Variable Estimates

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 9). 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.

Figure 9: 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.

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:

$$\text{Women}_{kt} = \psi \times \widehat{\text{Women}}_{kt} + \zeta_k + \omega_t + \xi_{kt} \tag{15}$$

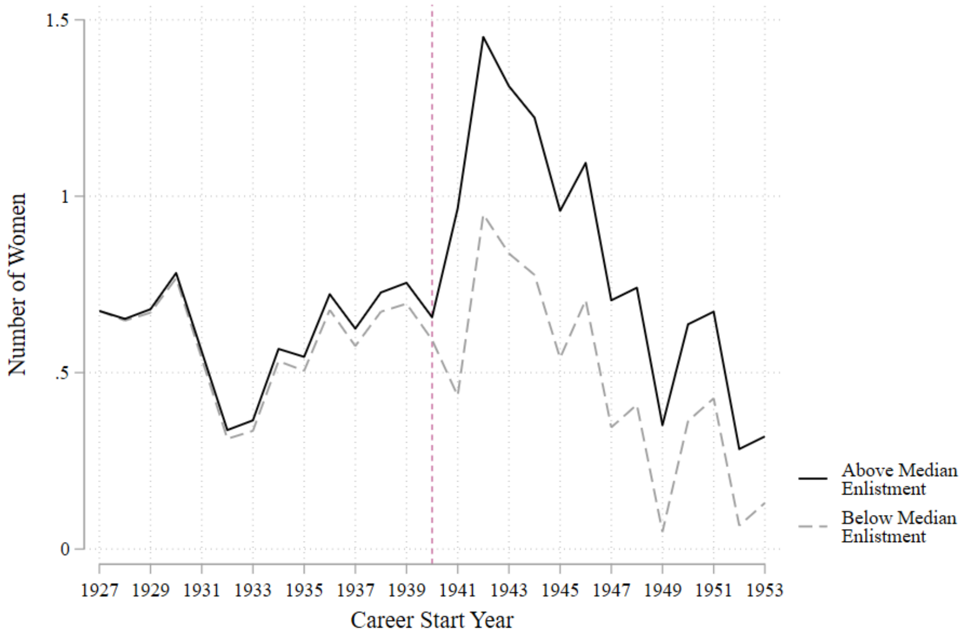
where the outcome variable Women_{kt} counts women entering field k in year t. Field fixed effects ζ_k capture time-invariant differences in female entry across fields, for instance, due to persistent norms and inflexible working conditions. Time fixed effects ω_t control for changes in female entry 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:

$$\widehat{\text{Women}}_{kt} = \widehat{\text{pre-war trend}}_{k|t<1940} + \gamma \times \text{enlistment}_k \times \mathbb{1}[t > 1939] \tag{16}$$

The parameter γ is the estimate of the causal effect of exposure to enlistment on female entry (from equation 14). The treatment variable enlistment_k is the share of male scientists in field k who enlisted and $\mathbb{1}[t > 1939]$ indicates years after the draft. The product of a field's enlistment rate, enlistment$_k$, the estimated proportional impact on female entry, γ, and the post-1939 dummy, $\mathbb{1}[t > 1939]$, measures 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.

In the physical sciences, more women enter science in fields with higher rates of enlistment after the draft (Figure 10).

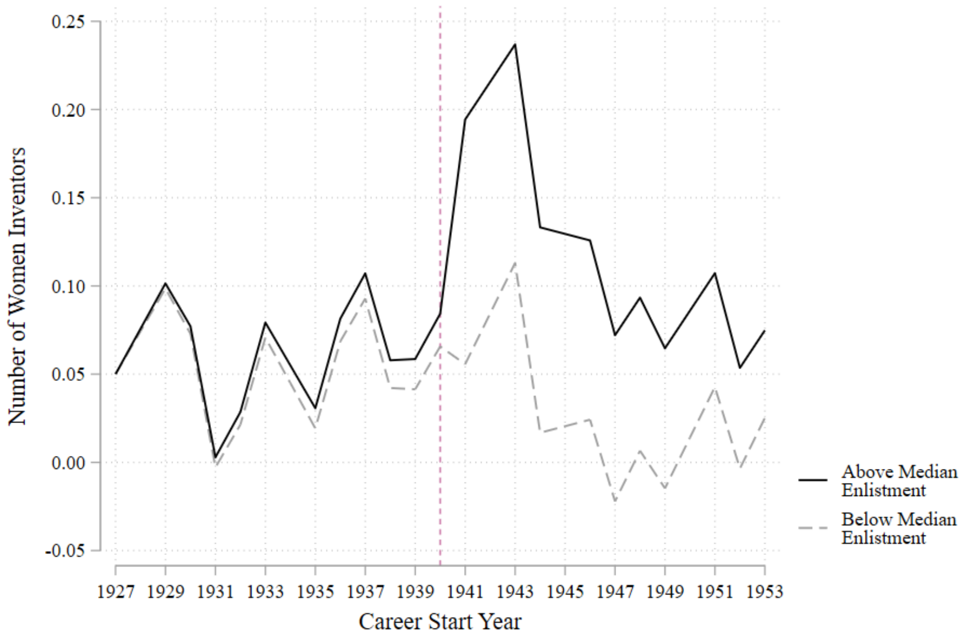
Figure 10: Entry of Female Scientists into the Physical Sciences



Note: Female scientists entering research fields in the physical sciences. Smoothed and centered averages.

The number of female inventors increases in fields with high exposure to enlistment relative to fields with low exposure (Figure 11). Until 1939, comparable numbers of female inventors enter science in fields with low and high exposure to enlistment. By 1943, the number of future female inventors entering science increases by a factor of 5 in fields with high exposure to enlistment and stays roughly constant in fields with low exposure.

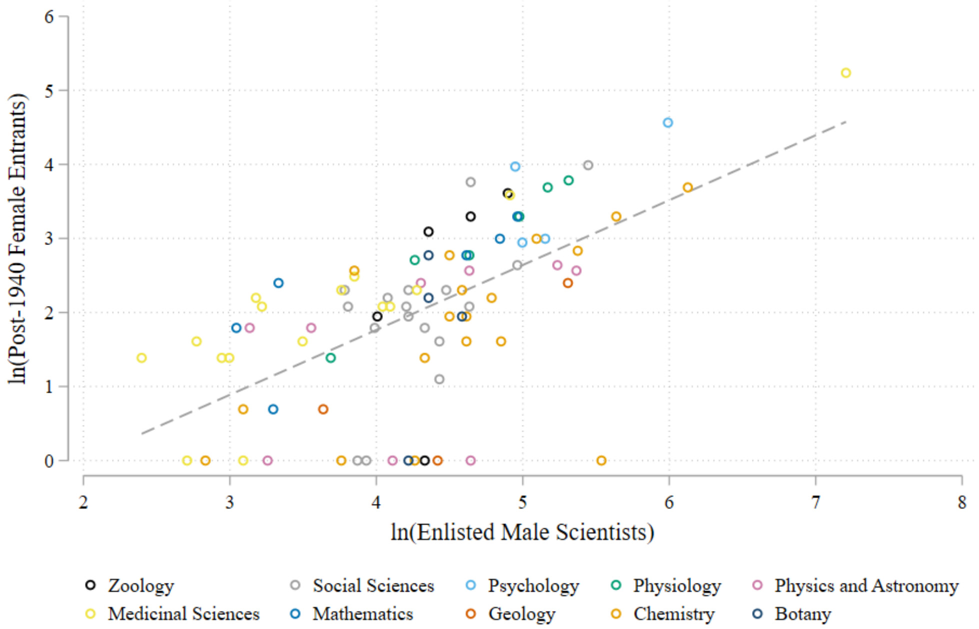
Figure 11: Entry of Female Inventors into the Physical Sciences



Note: Female inventors entering research fields in the physical sciences. Smoothed and centered averages.

Across all fields, the correlation between the natural logarithm of the number of enlisted men and the natural logarithm of women entering science is $\rho = 0.61$ (Figure 12).

Figure 12: Female Entry at the Field Level After the Draft



Note: Natural log of post-draft female entry against natural log of male army enlistments across research fields for draft-eligible birth cohorts (1897-1927). The correlation between the logged counts is $\rho = 0.61$. Fields are labeled by the median scientist's discipline.

Using 2SLS, we estimate the first stage (in equation 15) jointly with the second-stage equation:

$$Y_{kt} = \mu \times \text{Women}_{kt} + \eta_k + \theta_t + \varepsilon_{kt} \tag{17}$$

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 15). 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 an increase in the share of enlisted men in field k affects the number of inventors among female entrants in 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 3, 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 4, column 4): 1 additional woman becomes an inventor for every 5 women entering the physical sciences. In the biological sciences, 1 additional woman becomes an inventor for every 20 women.

OLS estimates are much smaller than 2SLS estimates, which suggests that OLS underestimates the true effect on female innovation. The OLS estimate for the physical sciences (0.095) is 51% smaller than the 2SLS coefficient (Table 3, column 1), and the OLS estimate for the biological sciences (0.039) is 21% smaller than the 2SLS estimate (Table 4, column 1). This 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 (Rossiter 1982, p.2) that are further removed from cutting-edge research and offer fewer opportunities to patent.

The field-level enlistment shock is a strong and meaningful instrument for female entry. Specifically, the predicted number of female entrants is strongly correlated with the number of women entering research fields after the draft (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 that compliance was higher in the biological than the physical sciences. In the biological sciences, 0.536 women join for every women predicted to join

(Table 4, column 3), compared with 0.175 in the physical sciences (Table 3, column 3). These results are consistent with stronger barriers to entry in the physical sciences.

Table 3: Field-Level IV Estimates for Female Scientists 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 research fields in the physical sciences. Standard errors clustered by research field. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Table 4: Field-Level IV Estimates for Female Scientists 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 research fields in the biological sciences. Standard errors clustered by research field. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

7. Individual-Level Instrumental Variable 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 fields for which they have more talent.

Scientist-level IV estimates allow us to estimate the effects of working in a field conditional on talent.[17] How much more likely is the same female scientist 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.[18] 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 a female scientist's probability to invent, we instrument her choice of research field with the field choices of all other women entering science in the same year. We estimate 2SLS regressions for all women in science:

$$\mathbb{1}[Inventor]_i = \beta_{iIV,phy} \times \mathbb{1}[\text{Phys. Sc.}]_i + \beta_{iIV,bio} \times \mathbb{1}[\text{Bio. Sc.}]_i + \theta_t + \varepsilon_i \tag{18}$$

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 and biological sciences; 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 (18) 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:

$$\mathbb{1}[\text{Phys. Sc.}]_i = \gamma_1 \times Z_{i,t}^{Phys.Sc.} + \gamma_2 \times Z_{i,t}^{Bio.Sc.} + \vartheta_t + \varepsilon_i \tag{19}$$

$$\mathbb{1}[\text{Bio. Sc.}]_i = \delta_1 \times Z_{i,t}^{Phys.Sc.} + \delta_2 \times Z_{i,t}^{Bio.Sc.} + \vartheta_t + \varepsilon_i \tag{20}$$

The research field that woman i enters is predicted by the leave-one-out average field choices

[17] While we cannot control for talent, exogenous variation at the individual level creates an assignment to fields that is as good as random. Quasi-random assignment solves the talent selection problem because it makes field choice independent of potential outcomes (Angrist and Pischke 2009, p.12).

[18] Rossiter (1982, p.2-4) notes that even when "efforts were being made to recruit women scientists, the positions offered to most of them were chiefly at the low levels." Hiring in science followed "the classic 'old-boy' recruitment patterns: the top men appointed men they knew already, and they in turn recruited their teams from among their own university and personal acquaintances."

$Z_{i,t}$ 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 in equation (18). We calculate the leave-one-out average field choices of individual i as

$$Z_{i,t}^{Phys.Sc.} = \frac{1}{N_t - 1} \sum_{j \neq i, t=t}^{J} \mathbb{1}[\text{Phys. Sc.}]_j \quad Z_{i,t}^{Bio.Sc.} = \frac{1}{N_t - 1} \sum_{j \neq i, t=t}^{J} \mathbb{1}[\text{Bio. Sc.}]_j \tag{21}$$

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.

The exclusion restriction for the IV estimates is that factors influencing the field choice of *other* female scientists entering a field in the biological of physical sciences in a given year do not influence the probability that a female scientist i entering the same field becomes an inventor at some stage of her career. If the exclusion restriction holds, 2SLS estimates identify the causal effect of entering the physical or biological sciences on the probability that female scientist i will become an inventor.

Overidentifying restriction tests support the 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 enlistments at the field-level as an additional push-factor that draw women into high-enlistment fields. To implement the overidentifying restriction tests, we include additional interaction terms between leave-one-out shares $Z_{i,t}$ and the field-level enlistment rate, creating two instruments per field. ψ_k denotes the share of enlisted scientists in field k:

$$\mathbb{1}[\text{Phys. Sc.}]_i = \gamma_1 Z_{i,t}^{Phys.Sc.} + \gamma_2 Z_{i,t}^{Bio.Sc.} + \gamma_3 Z_{i,t}^{Phys.Sc.} \times \psi_k + \gamma_4 Z_{i,t}^{Bio.Sc.} \times \psi_k + \vartheta_t + \varepsilon_i \tag{22}$$

$$\mathbb{1}[\text{Bio. Sc.}]_i = \delta_1 Z_{i,t}^{Phys.Sc.} + \delta_2 Z_{i,t}^{Bio.Sc.} + \delta_3 Z_{i,t}^{Phys.Sc.} \times \psi_k + \delta_4 Z_{i,t}^{Bio.Sc.} \times \psi_k + \vartheta_t + \varepsilon_i \tag{23}$$

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 6, columns 5), with a Hansen J test statistic of 2.208 and a p-value of 0.332, and therefore correctly excluded from equation (18).

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 percentage points (Table 5, column 5). Entering the biological sciences rather than social sciences increases a women's probability of becoming an inventor by 3.1 percentage points (Table 5, column 5).

IV estimates are only slightly larger than OLS estimates (at 11.4 and 3.0 pp. 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. Scientist-level IV estimates are robust to including controls for individual characteristics, including controls for PhD recipients and scientists working in industry (Tables A.4 and A.5).

Table 5: 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. Standard errors clustered by research field. Sample limited to female scientists. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 6: 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					4697.391
Kleibergen-Paap F					602.174
Hansen J					2.084
Hansen J (p-value)					0.353

Note: IV regressions estimated with 2SLS. Standard errors clustered by research field. Sample limited to female scientists. *** p < 0.01 ** p < 0.05 * p < 0.10

8. Selection

With gender distortions, the Roy model predicts that women face a higher talent threshold to enter fields with patenting than men. This result is consistent with a "Marie Curie strategy" (Rossiter 1982, p.159), requiring women to be better than men to get the same jobs:

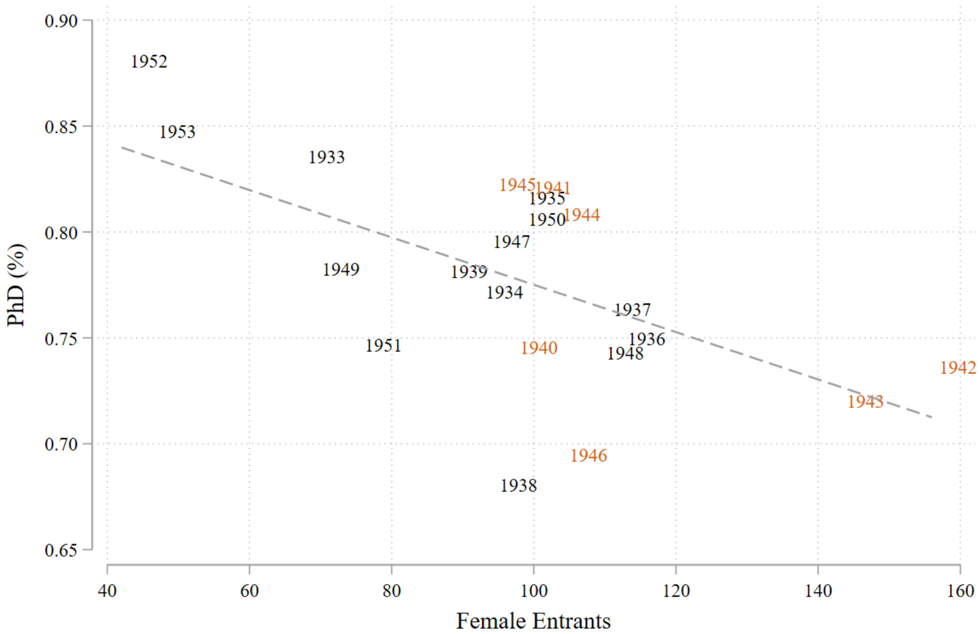
> "to be considered 'equal' to men, (...) women had to be 'better'. The most determined of them (the 'survivors') therefore internalized this double standard and settled down to becoming 'Madame Curies': earning an extra degree was just the first step in a lifelong struggle to overcompensate for being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

Consistent with positive selection, we find that female scientists are significantly more likely to hold a PhD than men: 79.8% of women have a doctorate, compared with 70.1% of men (p-value = 0.00). In the physical sciences, 81% of women have a PhD, compared with 69% of men (p-value = 0.00). Women are also significantly more likely to hold a degree from an elite university: 44% of female scientists have an Ivy Plus degree, compared with

35% of men (p-value = 0.00).[19] Female scientists are also significantly more likely to have earned degrees from universities that were considered elite in the first half of the 20th century (Slosson 1910; Thelin 2011, Table A.6).

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 1.1 percentage point lower (Figure 13). In the physical sciences, the share of women with a PhD in an entry cohort decreases by 4.9 percentage points for every ten additional female entrants (Figure 14).

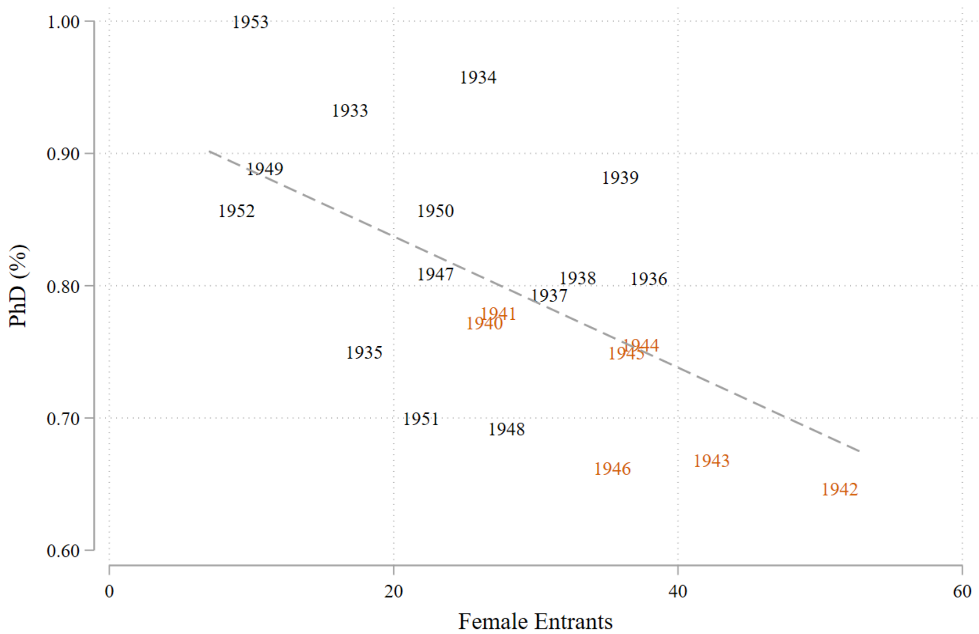
Figure 13: Share of Female Scientists with a PhD by Entry Cohort



Note: Share of female scientists with a PhD by entry cohort against the size of the female entry cohort (linear fit: -0.0011***). WWII entry cohorts (1940-46) highlighted.

[19]These differences hold in academia and industry. 83% of women in academia have a PhD (compared with 77% of men) and 39% have an Ivy+ degree (compared with 30% of men). In industry, 56% of women have a PhD (compared with 51% of men) and 32% of women hold an Ivy Plus degree (compared with 25% of men).

Figure 14: Share of Female Scientists with a PhD by Entry Cohort in the Physical Sciences



Note: Share of female scientists with a PhD in the physical sciences by entry cohort against the size of the female entry cohort (linear fit: -0.0049***). WWII entry cohorts (1940-46) highlighted.

If female scientists are positively selected they may create innovations that are better than those of male scientists. Using citations as a measure of quality, we find that patents by female scientists are significantly more likely to be among the most highly cited patents in their discipline (Table 7): female patents are 3.0 pp. more likely to be among the top 5 percent most cited patents in their discipline and year of patent issue, 4.4 pp. more likely to be in the top 10 percent, and 5.1 pp. more likely to be in the top 25 percent. On average, female patents receive 8.0% more citations than male patents (p = 0.12).[20]

[20] Notably, patents by female inventors tend to be cited less (Subramani and Saksena (2024), examining citations between for patents by male and female inventors between 1976 and 2023).

Table 7: Patent Forward Citations

Y =	(1) Top 5% Citations	(2) Top 10% Citations	(3) Top 25% Citations	(4) ln(Citations)
Female	0.030**	0.044**	0.051**	0.080
	(0.012)	(0.017)	(0.024)	(0.051)
Publication Year FE	✓	✓	✓	✓
Patent Class FE	✓	✓	✓	✓
Research Field FE	✓	✓	✓	✓
Mean Y	0.058	0.120	0.299	1.638
Observations	91,250	91,250	91,250	91,250
R-squared	0.015	0.020	0.029	0.082

Note: OLS estimates at the patent level for patents filed between 1920 and 1970. Top 5%, 10%, and 25% defined as binary indicators for all patents that receive at least as many forward citations as patents ranked at the respective percentile in their discipline within the year of patent issue. Sample limited to single-inventor patents and discipline-year cells in which at least 10 patents are filed. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

9. COUNTERFACTUALS

At the current rate of change, it will take 118 years to reach gender parity in invention (Bell et al. 2019).[21] What would happen to the persistent innovation gender gap if more women entered the physical and biological sciences? How many more women would become inventors? Would the innovation gender gap close more quickly? And would the total number of inventions rise or fall? To answer these questions, we present counterfactual estimates of innovation under the assumption that women would enter STEM at the same rate as men.

Counterfactual share of women among scientist-inventors

Trends towards gender parity for scientist-inventors born in the 1920s anticipate trends towards parity for inventors today (Figure 1). Among scientist-inventors born between 1921 and 1929, the share of women increases by 0.23 percentage points per year, compared with 0.27 percentage points for inventors born after 1940 (Bell et al. 2019), and 0.31 combining both series.

To estimate the counterfactual share of female inventors among scientist-inventors born 1921-29, we apply a simple non-parametric reweighting scheme (DiNardo et al. 1996), using the share of inventors among female scientists:[22]

$$\overline{\text{Inventor}}_g = f(k,g) \times Pr(\text{Inventor}|k,g) \tag{24}$$

[21] 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.

[22] DiNardo et al. (1996, p.11) produce a wage density for the counterfactual that unionization levels had remained constant between 1979 and 1988 "by replacing the relative weight of the densities of unionized vs. nonunionized workers of 1988 by that of 1979."

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 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 specification allows us to estimate counterfactual shares by – independently and jointly – varying the distribution of female scientists across fields and the probability to invent conditional on working in a field. Equation (24) allows us to estimate the counterfactual share of female inventors if women entered science in patent-intensive fields at the same rate as men. Specifically, we weigh the conditional inventor shares of women in k by the density $f'(k,g)$, 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 B8), and 2) the distribution of men across fields (Figure B9).

First, we estimate the effect of changing the distribution of female scientists $f(k,g)$ across fields – holding constant the probability $Pr(\text{Inventor}|k,g)$ that a female scientist has at least one patent. 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, $\overline{\text{Inventor}}'_g = 5.5\%$ of female scientists would be inventors,[23] 51.5% more than the observed share of 3.6%. We calculate counterfactual shares for each birth cohort by multiplying the observed share of female inventors with this relative increase. If women were distributed across fields similarly to men, 8.3% of female scientists would be inventors, 127.8% more than the observed share of 3.6%.

Next, we estimate counterfactuals allowing women entering science 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 a field on their probability to patent. Field-level IV estimates imply that 19.5% of women entering the physical instead of social sciences become inventors, 4.9% of women entering the biological sciences instead of the social sciences, and 14.6% of women entering the physical instead of the biological sciences become inventors. Individual-level IV estimates imply that 11.6% of women entering the physical rather than the social science become inventors, 3.0% of women entering the biological rather than social sciences, and 8.6% of women entering the physical rather than biological sciences.

Counterfactuals that use field-level IV estimates to predict rates of invention for female scientists entering the physical and biological sciences imply that 7.4% of female scientists would be inventors, 105.8% more than the observed share of 3.6%. Counterfactuals that use individual-level IV estimates imply that 5.9% of female scientists would be inventors, 62.4% more than the observed share of 3.6%.

[23] 29% of female scientists work in the physical sciences, 9.4% of them are inventors. 42% of female scientists work in the biological sciences, 2.0% of them are inventors. 29% of female scientists work in the social sciences, 0.2% of them are inventors. Holding inventor shares constant, we calculate the counterfactual share by weighing them by the distribution of men across sciences: 52% of men are in the physical sciences, 30% in the biological sciences, and 18% in the social sciences so that $0.52 \times 9.4\% + 0.30 \times 2.0\% + 0.18 \times 0.2\% = 5.5\%$.

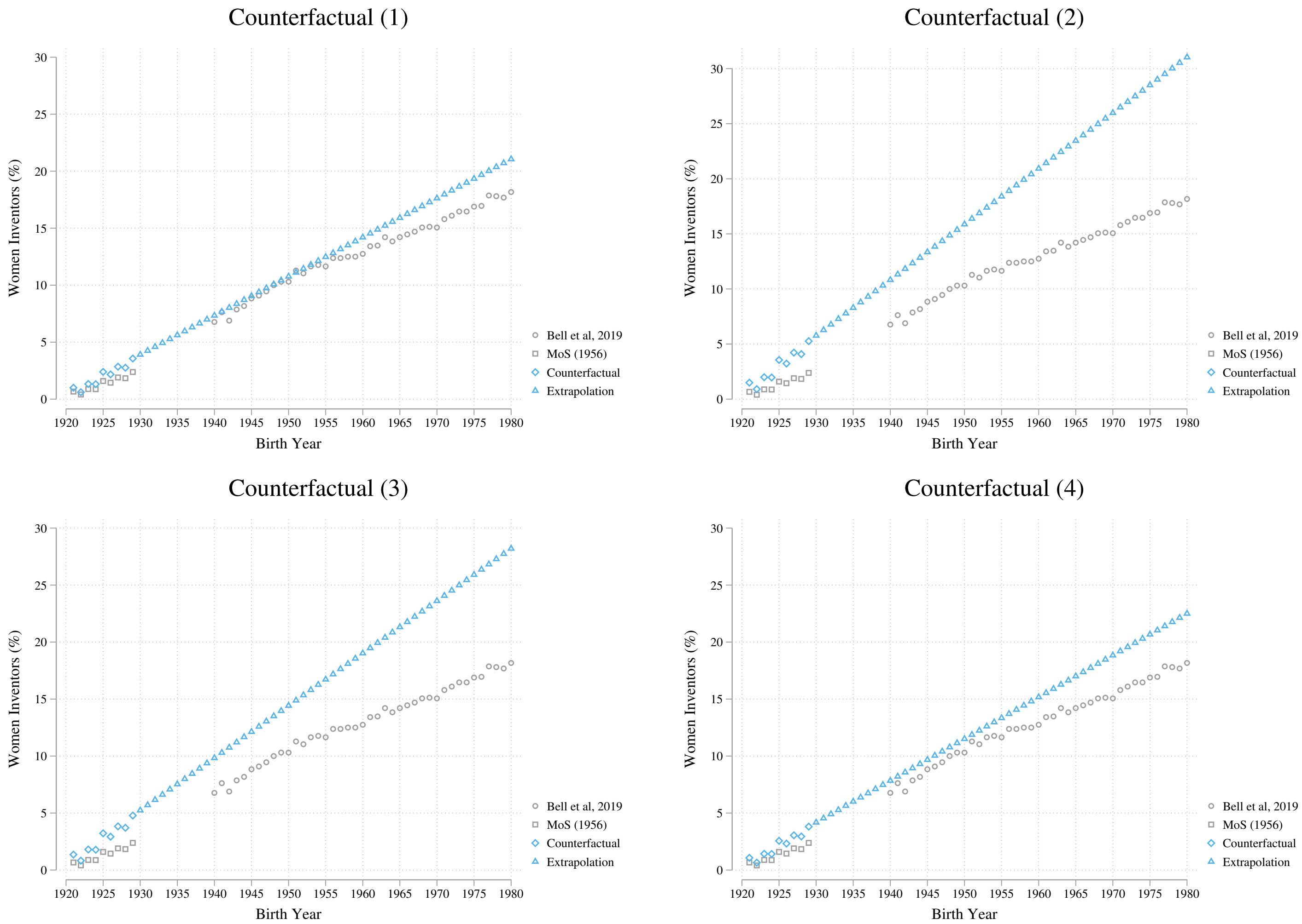
Counterfactual share of women among inventors born 1930-80

What would have been the share of women among inventors born 1930-80 had the share of female inventors grown at the rate implied by counterfactuals for scientist-inventors born 1921-29? And how long would it take to reach gender parity? To answer these questions, we extrapolate the counterfactual share of female inventors among scientists born 1921-29 to inventors born 1930-80 (Figure 1).

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.

Counterfactuals suggest that the share of female inventors today would be 15.9% to 70.8% higher had women entered STEM fields at the same rate as men (Figure 15). If women were as likely to work in the physical and biological sciences as men, 21.1% of inventors born in 1980 would have been women, 15.9% more than the observed share of 18.2% (counterfactual 1). If women were distributed across fields similarly to men, 31.0% of inventors would have been women, 70.8% more than the observed share of 18.2% (counterfactual 2). Using field-level IV estimates to predict rates of invention for women entering the physical and biological sciences, 28.2% of inventors would have been women, 55.2% more than the observed share of 18.2% (counterfactual 3). Using individual-level IV estimates, 22.5% of inventors would have been women, 23.9% more than the observed share of 18.2% (counterfactual 4).

Figure 15: Counterfactual Scenarios

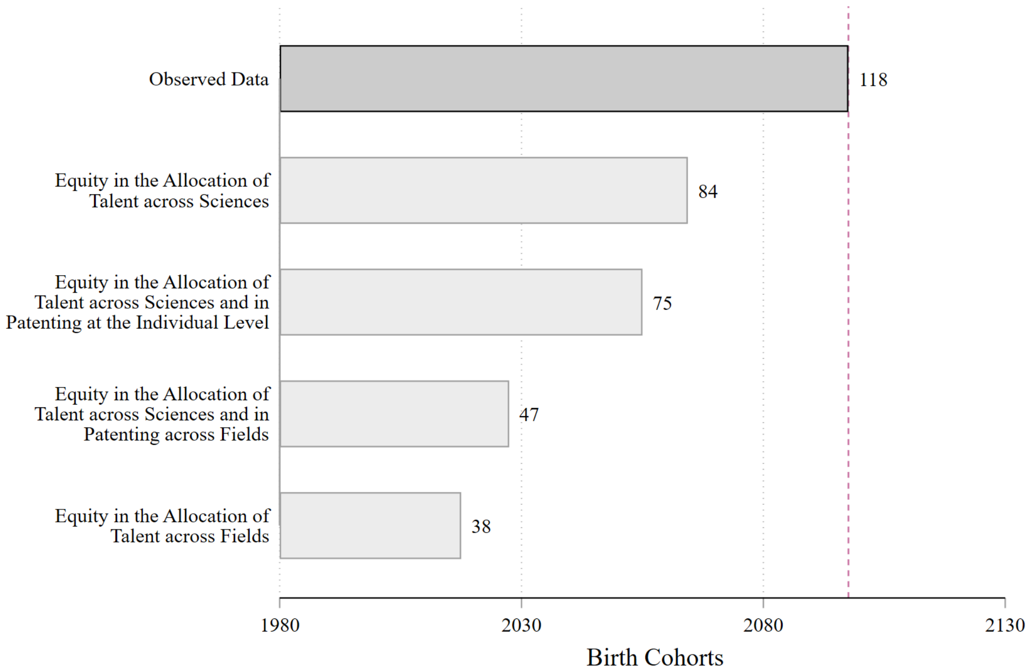


Note: Counterfactual (1) reweighs female scientists across sciences and keeps conditional patent rates constant. Counterfactual (2) reweighs female scientists across fields and keeps conditional patent rates constant. Counterfactual (3) reweighs female scientists across sciences and adjusts conditional patent rates using field-level IV estimates. Counterfactual (4) reweighs female scientists across sciences and adjusts conditional patent rates using individual-level IV estimates.

Counterfactual time to gender parity

Counterfactual estimates imply that improvements in the allocation of talent could help to close the innovation gender gap in 38 rather than 118 years, as implied by Bell et al. (2019) (Figure 16). If women were as likely to work in the physical and biological sciences as men (counterfactual 1), the first birth cohort to reach parity would be 2064 (34 years earlier). If women were distributed across fields like men (counterfactual 2), the first cohort to reach gender parity would be 2018 (80 years earlier). With equity in the allocation of talent across science and in patenting across fields (counterfactual 3), 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), the first cohort to reach gender parity would be 2055 (43 years earlier).

Figure 16: Counterfactual Timelines to Gender Parity



Note: Counterfactual timelines to gender parity in innovation. The dashed line marks the birth cohort of 2098, the first year with gender parity at observed rates of convergence from Bell et al. (2019).

Even lower-bound counterfactuals—allowing only for differences in levels of female inventors but not slopes (Figure B11)—imply significantly faster progress towards gender parity. If women had been as likely to enter STEM as men in the 1920s, but the share of women among inventors had continued to grow at the rate observed between 1940 and 1980, the first cohort to reach gender parity would be inventors born in 2075, 23 years earlier than implied by current estimates (Bell et al. 2019).

Counterfactual counts of inventors

Do female inventors displace male inventors or does female entry enlarge the overall inventor population? To answer this question, we decompose counterfactual entry:

$$r_{t,f'} \times (N_t + \Delta_t) = r_{t,f} N_t + s(r_{t,f'} - r_{t,f}) N_t + (1-s) \frac{(r_{t,f'} - r_{t,f})}{1 - r_{t,f'}} N_t \tag{25}$$

where $r_{t,f'}$ 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. Women replace men with substitution rate s. The right-hand side of equation (25) decomposes the counterfactual total number of female inventors into 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, 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.

The total increase in the number of inventors 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 $s = 0$. Then, the counterfactual share of women becomes

$$r_{t,f'} = \frac{r_{t,f} N_t + \Delta_t}{N_t + \Delta_t} \tag{26}$$

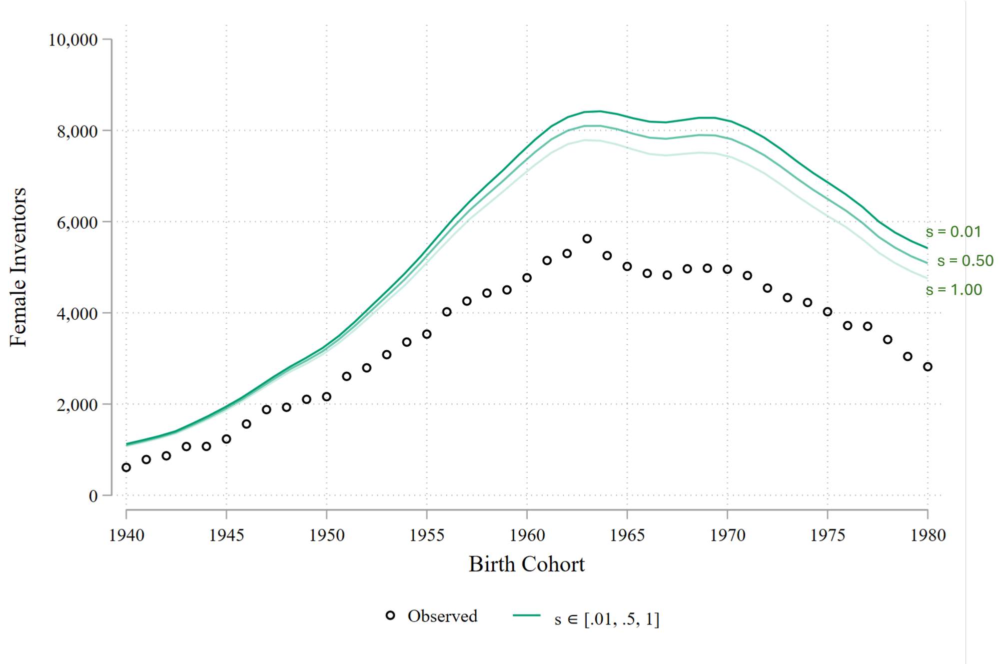
Solving for Δ_t yields

$$\Delta_t = \frac{(r_{t,f'} - r_{t,f}) N_t}{1 - r_{t,f'}} \tag{27}$$

$(N_t + \Delta_t)$ 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). A total of 1.17 million inventors were born in the US between 1940 and 1980.

Depending on the rate of substitution s, the number of female inventors increases by 47.7% to 60.5% (Figure 17; for counterfactual 3). If substitution is low (s = 0.01, so that every 100 female entrants replace 1 man), then the number of female inventors increases by 86,057 (60.5%) and the population of inventors increases by 85,196 (8.0%). If substitution is high (s = 1, so that every female entrant replaces a man), then the number of female inventors increases by 67,814 (47.7%) and the population of inventors remains unchanged.

Figure 17: Additional Female Inventors



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

Counterfactual counts of inventions

How many inventions would these additional female inventors create? To answer this question, we decompose the net gain in patents by inventors of birth cohort t under the counterfactual density f':

$$\text{Add. Patents}_{t,f'} = (r_{t,f'} - r_{t,f})N_t \left(\frac{1 - sr_{t,f'}}{1 - r_{t,f'}}\right) \times \text{Patents}_{t,w} - s(r_{t,f'} - r_{t,f})N_t \times \text{Patents}_{t,m} \quad (28)$$

where $(r_{t,f'} - r_{t,f})N_t \left(\frac{1-sr_{t,f'}}{1-r_{t,f'}}\right) \times \text{Patents}_{t,w}$, captures additional patents by female entrants at the substitution rate s. $s(r_{t,f'} - r_{t,f})N_t \times \text{Patents}_{t,m}$, 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.[24] Intuitively, the counterfactual change in patents reflects the difference in patents per person by male and female inventors. Since male inventors patent more, the net change in patenting will be negative if the substitution rate s is sufficiently high.[25]

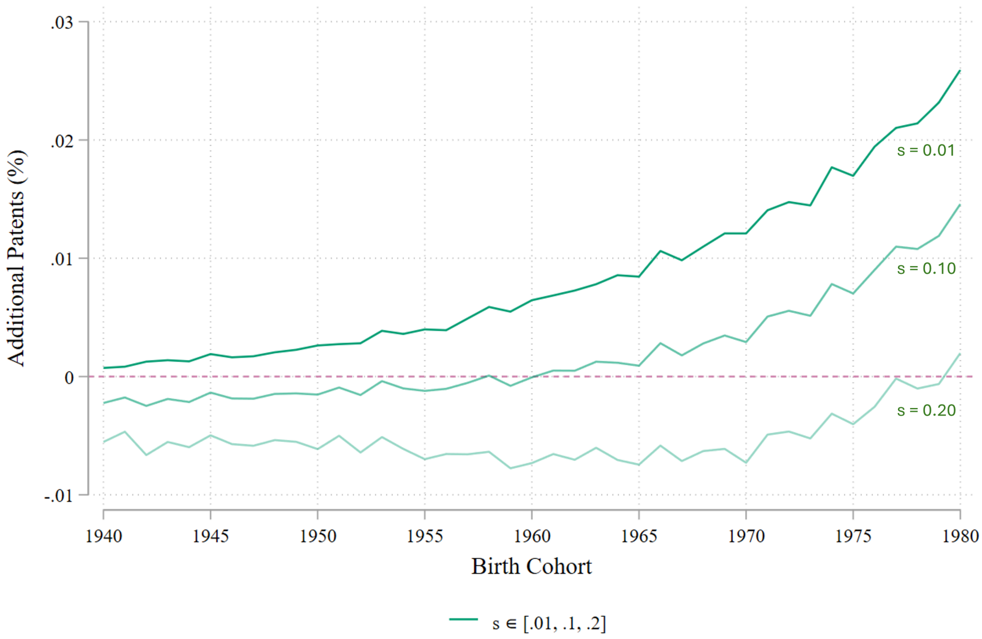
Depending on the rate of substitution s, the total number of patents increases up to 2.6% (Figure 18, for counterfactual 3, using field-level IV estimates) for different values for s. For inventors born in 1980, counterfactual patenting increases by 2.6% if substitution is low (s = 0.01), 0.2% more patents if substitution is moderate (s = 0.2) and reduces patenting if the

[24] Since patents per birth cohort by gender are only available for patents issued between 1996 and 2012 from Bell et al. (2019), we report changes as percentages to avoid censoring.

[25] Appendix Figure B10 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.

substitution rate is higher than 0.21.

Figure 18: Additional Inventions



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

10. 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. We present a Roy-model of field choice that shows that gender distortions impact both the participation and selection of women in invention.

Using the enlistment of male scientists 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. Field-level IV estimates imply that, for every 5 additional women entering the physical sciences, one additional woman would become an inventor. Counterfactual estimates imply that, had women entered STEM at the same rate as men, the innovation gender gap could close in 38 rather than 118 years, as implied by current estimates (Bell et al. 2019). We also find that 61% more American women born 1940-80 would have become inventors, and women would have produced roughly 3% more patents per birth cohort.

References

ACEMOGLU, DARON, DAVID H AUTOR, AND DAVID LYLE (2004): "Women, war, and wages: The effect of female labor supply on the wage structure at midcentury," *Journal of Political Economy*, 112 (3), 497–551.

ANEJA, ABHAY, OREN RESHEF, AND GAURI SUBRAMANI (2024): "Attrition and the Gender Patenting Gap," *Review of Economics and Statistics*, 1–31.

ANGRIST, JOSHUA D AND JÖRN-STEFFEN PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton university press.

BELL, ALEX, RAJ CHETTY, XAVIER JARAVEL, NEVIANA PETKOVA, AND JOHN VAN REENEN (2019): "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 134 (2), 647–713.

BERTRAND, MARIANNE, CLAUDIA GOLDIN, AND LAWRENCE F KATZ (2010): "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American Economic Journal: Applied Economics*, 2 (3), 228–255.

BIASI, BARBARA AND PETRA MOSER (2021): "Effects of copyrights on science: Evidence from the WWII book republication program," *American Economic Journal: Microeconomics*, 13 (4), 218–260.

BLINDER, ALAN S (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 436–455.

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

CARRELL, SCOTT E, MARIANNE E PAGE, AND JAMES E WEST (2010): "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 125 (3), 1101–1144.

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

CATTELL, J MCKEEN (1906): "A statistical study of American men of science: The selection of a group of one thousand scientific men," *Science*, 24 (621), 658–665.

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

CHETTY, RAJ, DAVID J DEMING, AND JOHN N FRIEDMAN (2023): "Diversifying society's leaders? The causal effects of admission to highly selective private colleges," *NBER Working Paper*.

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

COTTON, JEREMIAH (1988): "On the Decomposition of Wage Differentials," *The Review of Economics and Statistics*, 236–243.

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

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

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti (2004): "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 119 (4), 1249–1299.

Goldin, Claudia (1988): "Marriage bars: Discrimination against married women workers, 1920's to 1950's," *NBER Working Paper*.

——— (2014): "A grand gender convergence: Its last chapter," *American Economic Review*, 104 (4), 1091–1119.

Goldin, Claudia and Lawrence F Katz (2016): "A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation," *Journal of Labor Economics*, 34 (3), 705–746.

Goldin, Claudia and Claudia Olivetti (2013): "Shocking labor supply: A reassessment of the role of World War II on women's labor supply," *AEA Papers & Proceedings*, 103 (3), 257–262.

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

Gross, Daniel P and Bhaven N Sampat (2023): "America, jump-started: World War II R&D and the takeoff of the US innovation system," *American Economic Review*, 113 (12), 3323–3356.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow (2019): "The allocation of talent and US economic growth," *Econometrica*, 87 (5), 1439–1474.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger (2022): "Gender gaps in academia: Global evidence over the twentieth century," *Available at SSRN 4150221*.

Jaworski, Taylor (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 (1), 169–195.

Jensen, Kyle, Balázs Kovács, and Olav Sorenson (2018): "Gender differences in obtaining and maintaining patent rights," *Nature Biotechnology*, 36 (4), 307–309.

Kahn, Shulamit and Donna Ginther (2017): "Women and STEM," *NBER Working Paper*.

Kim, Scott Daewon and Petra Moser (2024): "Women in Science. Lessons from the Baby Boom," *Working Paper*.

Kitagawa, Evelyn M (1955): "Components of a Difference Between Two Rates," *Journal of the American Statistical Association*, 50 (272), 1168–1194.

MINCER, JACOB (1978): “Family migration decisions,” *Journal of Political Economy*, 86 (5), 749–773.

MOSER, PETRA AND SAHAR PARSA (2024): “McCarthy and the Red-ucators: Effects of Political Persecution on Science,” *Working Paper*.

MOSER, PETRA, SAHAR PARSA, AND SHMUEL SAN (2024): “Immigration, Science, and Invention. Lessons from the Quota Acts,” *Working Paper*.

NGUYEN, BANG AND PETRA MOSER (2024): “Operation Paperclip: Nazi Scientists and US Innovation,” *Working Paper*.

OAXACA, RONALD (1973): “Male-Female Wage Differentials in Urban Labor Markets,” *International Economic Review*, 693–709.

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

PAIROLERO, NICHOLAS, ANDREW TOOLE, CHARLES DEGRAZIA, MIKE HORIA TEODORESCU, AND PETER-ANTHONY 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, ARPITA, MATTHEW WISWALL, AND BASIT ZAFAR (2021): “College majors 1,” *The Routledge handbook of the economics of education*, 415–457.

ROSE, EVAN 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 (3), 673–711.

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

ROY, ANDREW DONALD (1951): “Some thoughts on the distribution of earnings,” *Oxford Economic Papers*, 3 (2), 135–146.

SARSONS, HEATHER, KLARITA GËRXHANI, ERNESTO REUBEN, AND ARTHUR SCHRAM (2021): “Gender differences in recognition for group work,” *Journal of Political Economy*, 129 (1), 101–147.

SINHA, ARNAB, ZHIHONG SHEN, YANG SONG, HAO MA, DARRIN EIDE, BO-JUNE HSU, AND KUANSAN WANG (2015): “An overview of microsoft academic service (mas) and applications,” in *Proceedings of the 24th international conference on world wide web*, 243–246.

SLOSSON, EDWIN EMERY (1910): *Great American Universities*, Macmillan.

SUBRAMANI, GAURI AND MICHELLE SAKSENA (2024): “Untapped Potential: Investigating Gender Disparities in Patent Citations,” *USPTO Economic Working Paper*.

THELIN, JOHN R (2011): *A history of American higher education*, JHU Press.

WASSERMAN, MELANIE (2023): “Hours constraints, occupational choice, and gender: Evidence from medical residents,” *The Review of Economic Studies*, 90 (3), 1535–1568.

WISWALL, MATTHEW AND BASIT ZAFAR (2018): "Preference for the workplace, investment in human capital, and gender," *The Quarterly Journal of Economics*, 133 (1), 457–507.

A. ADDITIONAL TABLES

Table A.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	***
PhD	0.707	0.455	0.701	0.458	0.798	0.401	0.097	***
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 A.2: 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.3: OLS Estimates of the Innovation Gender Gap with Alternative Measures of Publications

	(1) Publications	(2) ln(Publications + 1)	(3) ln(Publications + 0.1)	(4) IHS(Publications)
Female	-0.192***	-0.163***	-0.170***	-0.159***
	(0.004)	(0.005)	(0.003)	(0.005)
Pub. Count	0.000***			
	(0.000)			
Publications × Female	-0.000			
	(0.000)			
ln(Publications + 1)		0.023***	0.018***	
		(0.001)	(0.001)	
ln(Publications + 1) × Female		-0.012***	-0.011***	
		(0.003)	(0.001)	
IHS(Publications)				0.021***
				(0.001)
IHS(Publications) × Female				-0.011***
				(0.002)
Constant	0.226***	0.183***	0.200***	0.177***
	(0.002)	(0.003)	(0.002)	(0.003)
Birth year FE	✓	✓	✓	✓
Birth state FE	✓	✓	✓	✓
Mean Y	0.221	0.221	0.221	0.221
Observations	70,033	70,033	70,033	70,033
R-squared	0.022	0.026	0.028	0.027

Note: Outcome variable equal to one if a scientist obtains at least one patent during their career, zero otherwise. Publications measured as publication count (column 1), the natural logarithm of publications + 1 (column 2), the natural logarithm of publications + 0.1 (column 3), and the inverse hyperbolic sine of publications (column 4). Robust errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Table A.4: 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: 2SLS estimates. Inventor is an indicator scientists with at least one patent. Physical and Biological Sciences are indicators for scientists in the respective sciences. PhD is an indicator for scientists with a doctoral degree. Industry is an indicator for industry employment. Sample limited to female scientists. Robust errors clustered by field. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table A.5: 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					4640.832
Kleibergen-Paap F					578.429
Hansen J					1.914
Hansen J (p-value)					0.384

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 a research field. Robust errors clustered by field in parentheses. *** p < 0.01 ** p < 0.05 * p < 0.10

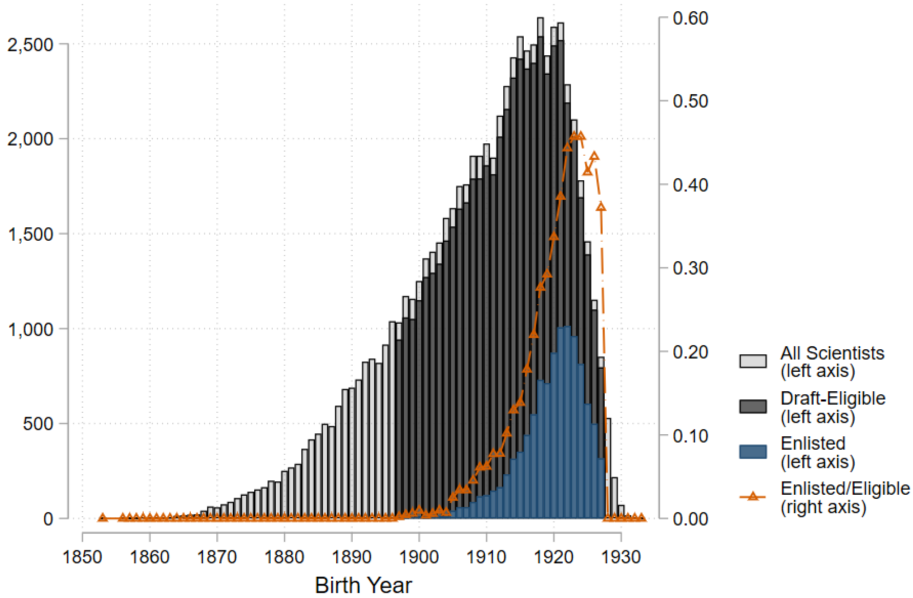
Table A.6: Elite University Degrees as a Measure of Selection

	Top 5 Universities in 1890 (Thelin 2011)	Ivy Plus (Chetty, Deming, and Friedman 2023)	Great Universities in 1910 (Slosson 1910)
Male Scientists	0.217	0.346	0.554
Female Scientists	0.290	0.438	0.630
Difference	-0.073***	-0.092***	-0.076***
Universities	Columbia, Cornell, Harvard, Johns Hopkins, Clark	Brown, Columbia, Cornell, Dartmouth, Duke, Harvard, MIT, Princeton, Stanford, Chicago, Penn, Yale	Berkeley, Chicago, Columbia, Cornell, Harvard, Illinois, Johns Hopkins, Michigan, Minnesota, Penn, Princeton, Stanford, Wisconsin, Yale

Note: Average elite university attendance for male and female scientists. Elite university attendance coded as 1 if a scientist earned an undergraduate, master's, or PhD degree from an elite university. Honorary degrees, fellowships, and scholarships omitted.

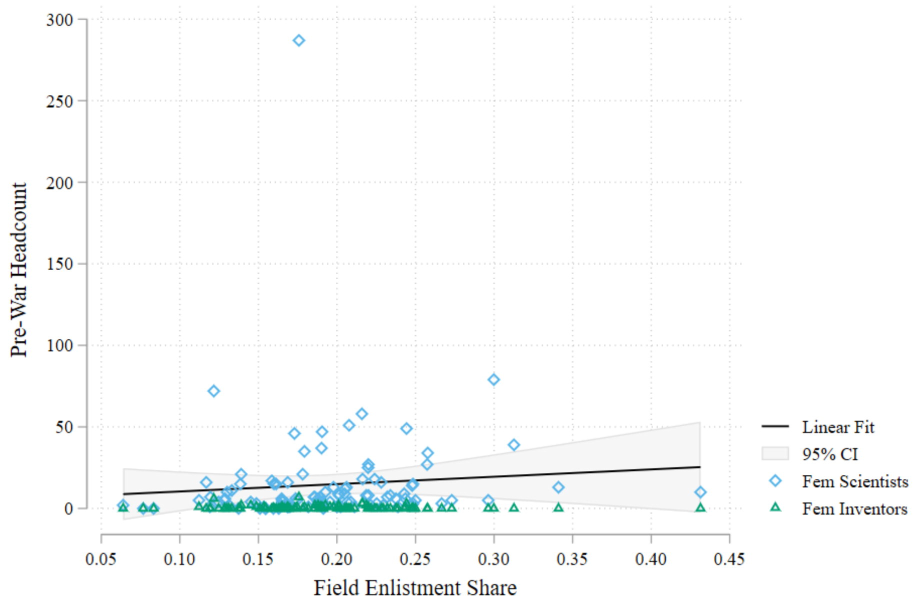
B. Additional Figures

Figure B1: Enlistment across Birth Cohorts



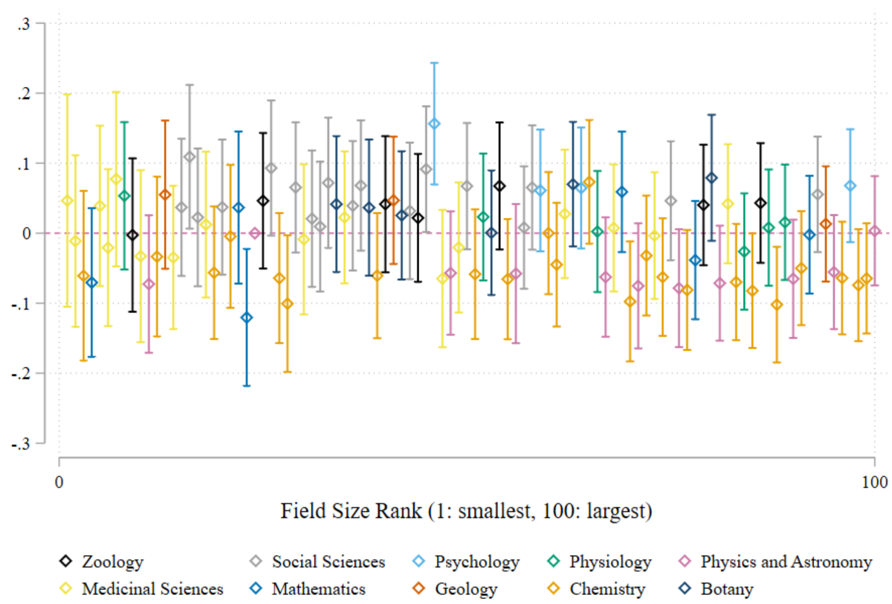
Note: Draft-eligible scientists defined as men born 1897-1927.

Figure B2: Correlation between Enlistment Share and Pre-1940 Female Entry



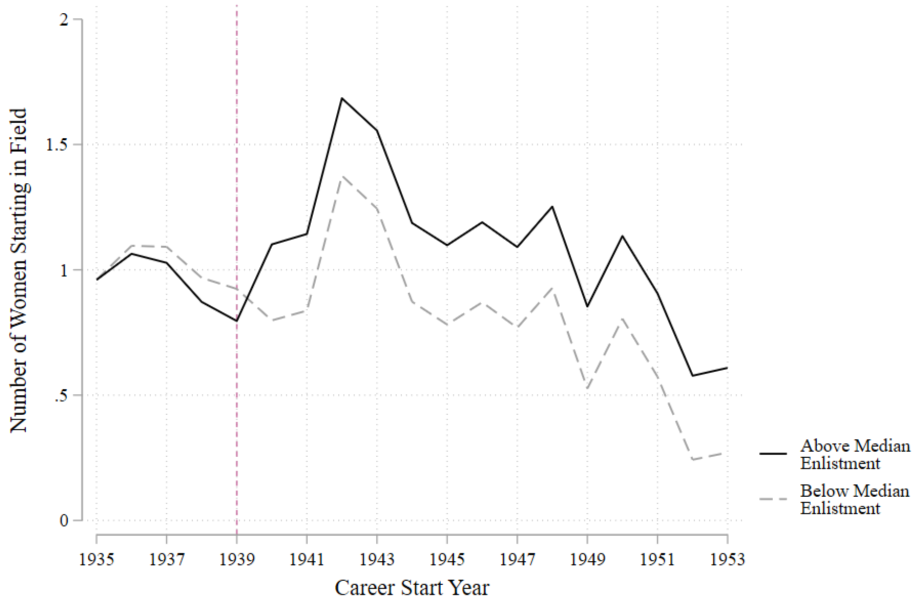
Note: Y-axis: Count of female scientists and female inventors per field before the WWII draft. X-axis: Field-level enlistment share. Linear fit: predicted values for the number of female scientists from a linear regression of female scientists on field-level enlistment rates. The correlation coefficient between female scientists and enlistment is 0.12, the correlation coefficient between female inventors and enlistment is -0.02.

Figure B3: 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. Among 100 estimated coefficients, 6 are significant at 5% and 1 is significant at 1%.

Figure B4: Mean Female Entry at Field-Level

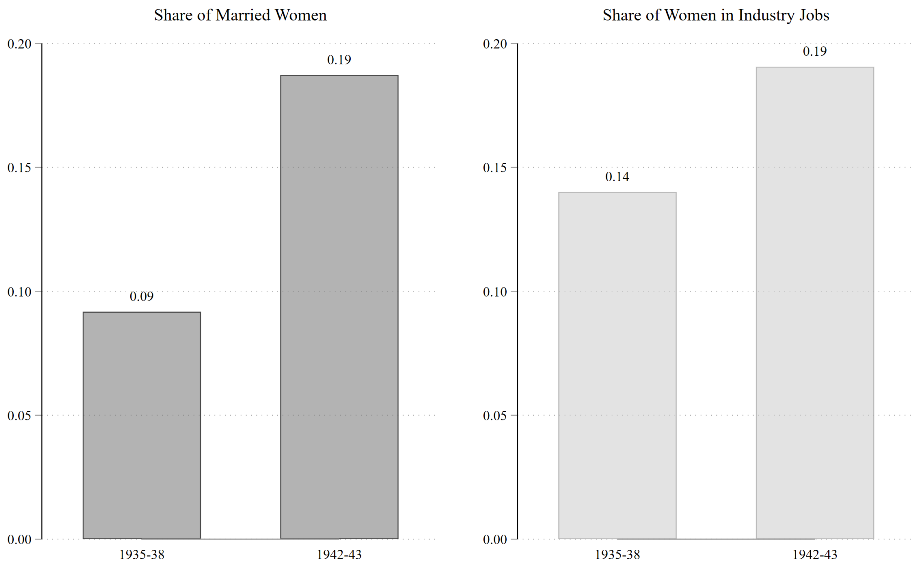


Note: Smoothed and centered average female entry in fields with above and below median enlistment among male scientists.

Figure B5: Age Distribution of Female Entrants, 1935-38 and 1942-43

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.

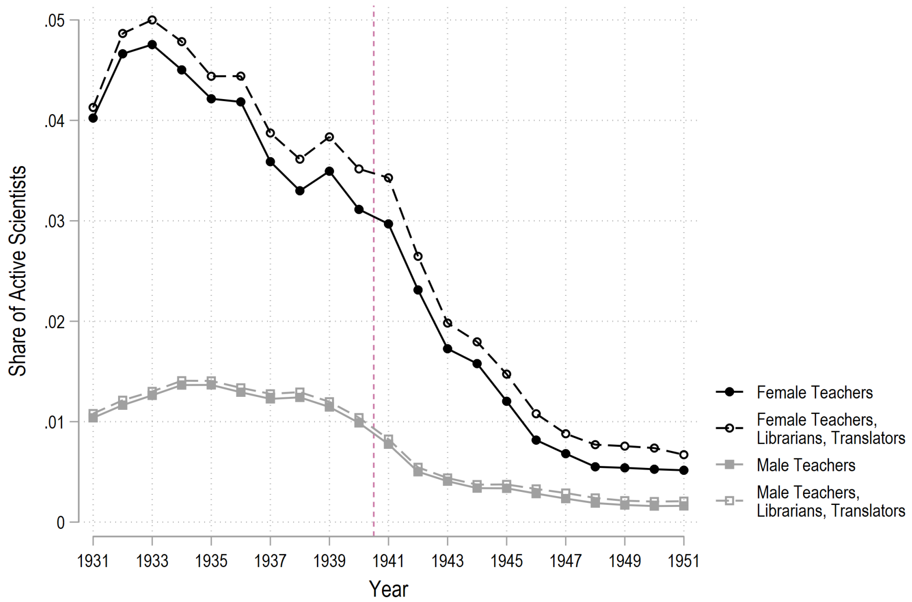
Figure B6: Share of Married Women and Women in Industry Jobs, 1935-38 and 1942-43



N = 713

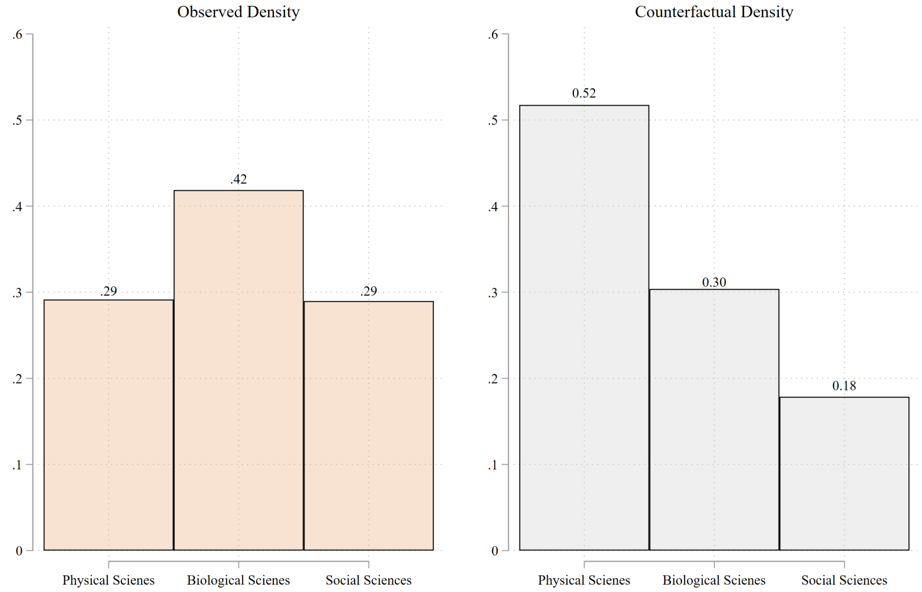
Note: The left figure plots the share of married women among female entrants before WWII (1935-38) and during WWII (1942-43). We measure marital status in the year of entry. The right figure shows the share of women working in industry jobs among women entering science before (1935-38) and during WWII (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 B7: Share of Scientists Working as Teachers, Librarians, and Translators



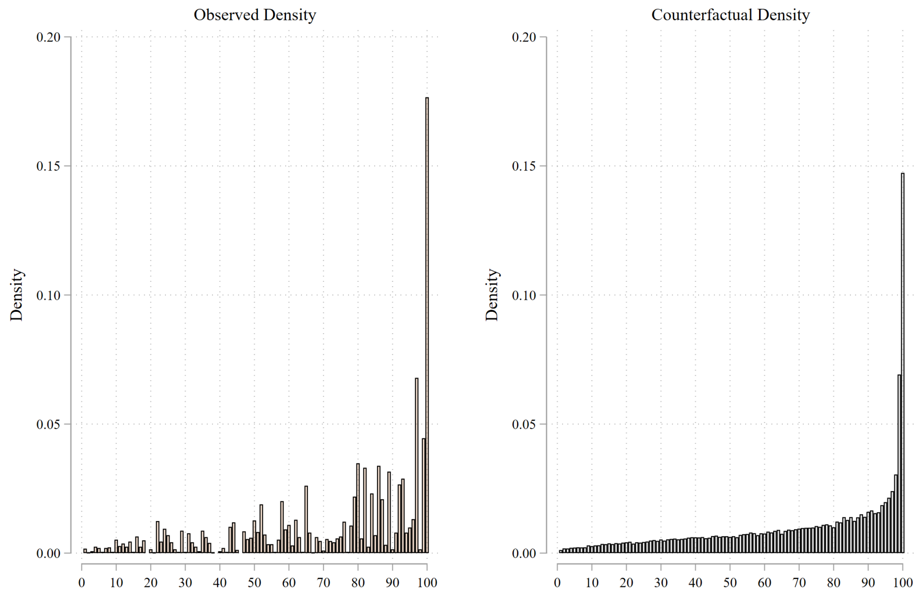
Note: Share of active scientists who work as teachers, librarians, or translators per year. Occupations classified based on job titles in MoS (1956). Year denotes when scientists worked in an occupation.

Figure B8: Observed and Counterfactual Densities: Sciences



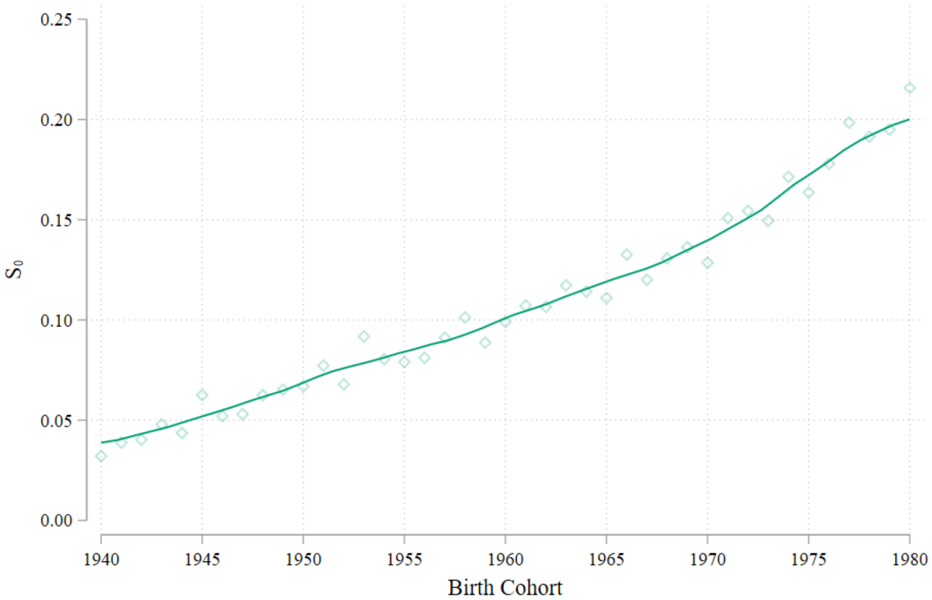
Note: Observed distribution of women (left) and men (right) across physical, biological, and social sciences in the MoS (1956). Reweighting counterfactuals use the distribution of men across sciences as a counterfactual for the distribution of women.

Figure B9: Observed and Counterfactual Densities: Fields



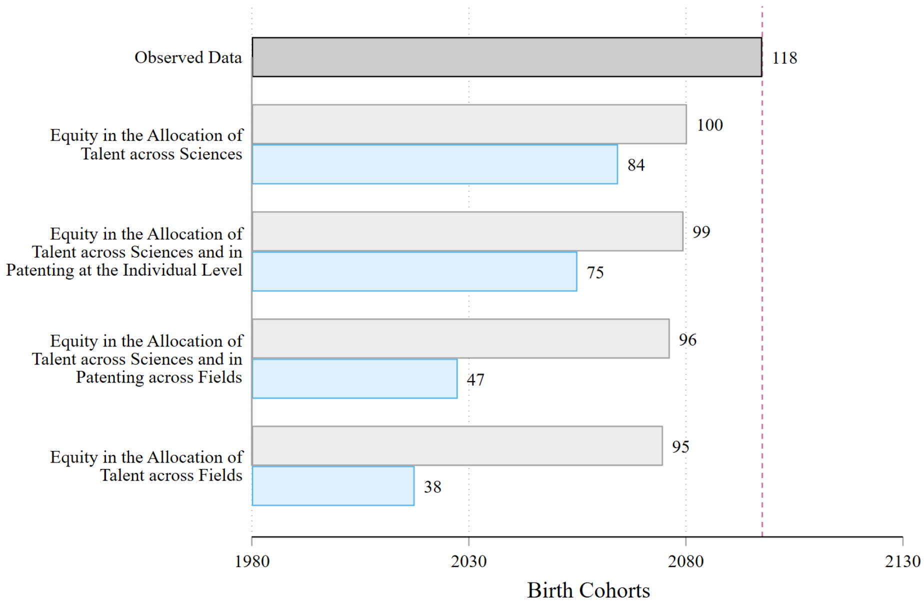
Note: Observed distribution of women (left) and men (right) across fields in the MoS (1956). Reweighting counterfactuals use the distribution of men across sciences as a counterfactual for the distribution of women. Fields ordered by the total number of scientists active in each field.

Figure B10: Breakeven Substitution Rate



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

Figure B11: Upper- and Lower-Bound Counterfactual Timelines to Gender Parity



Note: Counterfactual timelines to gender parity in innovation. Observed data from Bell et al. (2019). Upper-bound based on counterfactual level and slope, lower-bound on counterfactual level and observed slope.

C. Model Appendix

We rewrite Equations (6) - (12) using the full notation for cost and wage functions:

$$\bar{P}_w - \omega\left(\frac{1-\bar{P}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{P}_m^2}{2}\right) - (1-\bar{P}_w)^2 - d = \bar{S}_w - \omega\left(\frac{1-\bar{S}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{S}_m^2}{2}\right) - (1-\bar{S}_w)^2 \tag{29}$$

$$\bar{P}_w - \omega\left(\frac{1-\bar{P}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{P}_m^2}{2}\right) - (1-\bar{P}_w)^2 - d = 0 \tag{30}$$

$$\bar{S}_w - \omega\left(\frac{1-\bar{S}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{S}_m^2}{2}\right) - (1-\bar{S}_w)^2 = 0 \tag{31}$$

$$\bar{P}_m - \omega\left(\frac{1-\bar{P}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{P}_m^2}{2}\right) - (1-\bar{P}_m)^2 = \bar{S}_m - \omega\left(\frac{1-\bar{S}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{S}_m^2}{2}\right) - (1-\bar{S}_m)^2 \tag{32}$$

$$\bar{P}_m - \omega\left(\frac{1-\bar{P}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{P}_m^2}{2}\right) - (1-\bar{P}_m)^2 = 0 \tag{33}$$

$$\bar{S}_m - \omega\left(\frac{1-\bar{S}_w^2}{2}\right) - (1-\omega)\left(\frac{1-\bar{S}_m^2}{2}\right) - (1-\bar{S}_m)^2 = 0 \tag{34}$$

$$\frac{\frac{1-\bar{P}_w^2}{2}}{\frac{1-\bar{P}_m^2}{2}} = r \tag{35}$$

For $\omega = \frac{1}{2}$ this system of equations can be simplified:

$$\bar{S}_m^2 = 6 + 3(-4+\bar{S}_w)\bar{S}_w \tag{36}$$

$$6 + 3(-4+\bar{S}_m)\bar{S}_m = \bar{S}_w^2 \tag{37}$$

$$4d + \bar{S}_m^2 + 3(-4+\bar{P}_w)\bar{P}_w = 3(-4+\bar{S}_w)\bar{S}_w + \bar{P}_m^2 \tag{38}$$

$$3(-4+\bar{S}_m)\bar{S}_m + \bar{P}_w^2 = \bar{S}_w^2 + 3(-4+\bar{P}_m)\bar{P}_m \tag{39}$$

$$\frac{-1+\bar{P}_w^2}{-1+\bar{P}_m^2} = r \tag{40}$$

First, we solve for the talent thresholds in fields without patenting, $\bar{S}_w$ and $\bar{S}_m$, by eliminating $\bar{P}_w$, $\bar{P}_m$, r, and d:

$$3\bar{S}_m = 6 - 9\bar{S}_w + 2\bar{S}_w^2 \tag{41}$$

$$39\bar{S}_w^2 - 18\bar{S}_w^3 + 2\bar{S}_w^4 = 9 \tag{42}$$

For $\{\bar{S}_w, \bar{S}_m\} \in [0,1]$, the support of the talent distribution defined in the model, these equations have a unique solution at $\bar{S}_w = 3-\sqrt{6}$ and $\bar{S}_w = \bar{S}_m$.
Next, we solve for the talent thresholds in fields with patenting, $\bar{P}_w$ and $\bar{P}_m$, and the distortion parameter d. Eliminating $\bar{S}_w$ and $\bar{S}_m$ yields:

$$6+d-9\bar{P}_m+2\bar{P}_m^2-3\bar{P}_w=0 \tag{43}$$

$$6+3d-3\bar{P}_m-9\bar{P}_w+2\bar{P}_w^2=0 \tag{44}$$

$$-8-3d+3\bar{P}_m+9\bar{P}_w+8r+dr-9\bar{P}_m r-3\bar{P}_w r=0 \tag{45}$$

Since this system of equations is under-determined with four unknowns and three equations, we solve for these quantities as functions of the innovation gender gap r. Imposing $\{\bar{P}_w, \bar{P}_m\} \in [0,1]$, we obtain parameter expressions for $\bar{P}_w$, $\bar{P}_m$, and d:

$$\bar{P}_m = \frac{-6+\sqrt{21+2r+r^2}}{-3+r}$$

$$\bar{P}_w = \frac{\sqrt{3}\sqrt{-1+3r+\frac{24r}{-3+r}-\frac{4r\sqrt{21+2r+r^2}}{-3+r}}}{\sqrt{-3+r}}$$

$$d = \frac{-8-\frac{18}{-3+r}+8r+\frac{54r}{-3+r}+\frac{3\sqrt{21+2r+r^2}}{-3+r}-\frac{9r\sqrt{21+2r+r^2}}{-3+r}+\frac{9\sqrt{3}\sqrt{-1+3r+\frac{24r}{-3+r}-\frac{4r\sqrt{21+2r+r^2}}{-3+r}}}{\sqrt{-3+r}}-\frac{3\sqrt{3}r\sqrt{-1+3r+\frac{24r}{-3+r}-\frac{4r\sqrt{21+2r+r^2}}{-3+r}}}{\sqrt{-3+r}}}{3-r}$$

To maintain tractability, we define:

$$\theta = \frac{\sqrt{3}\sqrt{-1+3r+\frac{24r}{-3+r}-\frac{4r\sqrt{21+2r+r^2}}{-3+r}}}{\sqrt{-3+r}} \qquad \gamma = \sqrt{21+2r+r^2}$$

Substituting θ and γ:

$$\bar{P}_m = \frac{\gamma-6}{r-3} \qquad \bar{P}_w = \theta$$

$$d = \frac{-8-\frac{18}{-3+r}+\frac{3\gamma}{-3+r}+8r+\frac{54r}{-3+r}-\frac{9\gamma r}{-3+r}+9\theta-3r\theta}{3-r}$$

Further simplifying yields:

$$\bar{P}_m = \frac{\gamma-6}{r-3} \qquad \bar{P}_w = \theta$$

$$d = 3\theta - \frac{6+\gamma(3-9r)+22r+8r^2}{(r-3)^2}$$