

Gender Gaps in Academia: Global Evidence over the Twentieth Century *

Alessandro Iaria (University of Bristol)

Carlo Schwarz (Bocconi University)

Fabian Waldinger (Ludwig-Maximilians-Universität München)

May 15, 2026

Abstract

We study the evolution of gender gaps in academia over the 20th century. We assemble the most comprehensive global database of academics to date, covering scholars from every discipline across more than 150 countries and 7,000 universities. This unprecedented scope allows us to move beyond country- or discipline-specific case studies and analyze gender gaps in academic careers on a global scale. Four key findings emerge. First, women were systematically less likely to be hired into academic positions, with especially large gaps at prestigious universities, in STEM fields, and among full professors. Second, women published fewer papers than men, and we document a U-shaped relationship between gender gaps in hiring and publishing, indicating that barriers to entry and subsequent research output are intertwined. Third, women received fewer citations. Using a novel machine-learning approach, we show that this citation gap cannot be explained by women specializing in less-cited research areas. Fourth, women were less likely to be promoted to full professor, even conditional on publication and citation records.

*We thank the editor Oriana Bandiera, three knowledgeable referees, as well as Ran Abramitzky, David Card, Amy Finkelstein, Ben Jones, Muriel Niederle, Claudia Olivetti, Nina Roussille, and seminar participants at Berkeley, Bocconi, Bonn, Boston University, Boulder, Bristol, Florida, Frankfurt, Imperial College London, Munich, Nürnberg, NYU, Stanford, Turin, UCLA, Warwick, York, SOLE 2023, CEPR 2022, 2023, UCL-Warwick Women in Science Workshop 2023, and Barcelona Summer Forum 2022 for insightful comments and suggestions. We also thank Matthias Bing, Zishan Bhatti, Angela Buensuceso, Adam Brezinski, Le Van Cao, Giedrius Daubaris, Thomas Decker, Cerensu Dertli, Johannes Dieminger, Tomas Dimitrov, Leila Essa, Felicitas Filsinger, Nora Fingado, Lukas Franz, David Full, Franziska Ganz, Adrian Gashi, David Geiger, Julia Geltinger, Peter Heinrich, Robert Johannes, Simon Kallmaier, Katharina Knuth, Christoph Kuehn, Paulina Maier, Axel-Konstantin Marko, Lukas Mergele, Luisa Neumeier, Tobias Nowacki, Adrian Nutzinger, Franz Oertel, Julia Ostendorf, Nicolas Porschke, Joseph Prestwich, Felix Radde, Akos Reitz, Tabea Rudolf, Riyan Sawalha, Constanze Schwarz, Marie Spörk, Jan Stavrev, Nils Süßenbach, Paul von Tein, Luca Veh, Simon Velten, Constanze Vorrath, and Stefan Wies for excellent research assistance. Special thanks go to our data managers, Lena Greska and Sebastian Hager, who have been instrumental for the success of the data collection. We are also grateful to Martin Watzinger for sharing the Microsoft Academic Graph data. Waldinger gratefully acknowledges startup funding from ERC Starting Grant N°335573.

The production of ideas is central to economic growth in modern economies (e.g., Romer, 1986, 1990; Mokyr, 1992; Jones, 1995). Societies that generate more scientific and technological ideas expand the frontier of production and grow faster. Yet for most of modern history, the institutions responsible for producing ideas drew talent only from a subset of the population (Hsieh et al., 2019). Women, in particular, faced significant barriers or were entirely excluded from the idea-producing sector and other high-skilled professions. Even today, women account for just 9.5 percent of patents in OECD countries (OECD 2021).

These barriers were especially consequential in academia, one of the central institutions of idea production. Women were often prevented from pursuing academic careers, denied access to laboratories and professional networks, and passed over for senior positions. The case of Marie Curie, one of the most celebrated scientists in history, illustrates these barriers. Even after she won her first Nobel Prize in 1903, the Sorbonne refused to appoint her to a professorship. Only in 1908, after the tragic death of her husband Pierre and two years before her second Nobel Prize, did she become the first female full professor in the university’s history. Curie’s case is remembered precisely because she was extraordinary enough to overcome barriers that excluded many others. Her experience raises a broader question: how much scientific talent was lost because academic institutions failed to hire, support, recognize, and promote women?

A rich literature documents gender gaps in academic careers, examining differences in publications, citations, and recognition. Yet almost all of this work focuses on a single discipline, a single country, or a relatively short time horizon. As a consequence, important questions about academia as a global institution remain unanswered. In particular, we know little about the long-run evolution of gender gaps across countries and fields, and about how gender differences vary and interact across stages of academic careers.

Studying these questions has been hampered by a lack of data on women in academia. In this paper, we present the first comprehensive analysis of gender gaps in academia on a global scale, covering the entire 20th century. We trace the evolution of gender gaps across four key stages of academics’ careers: hiring, publishing, citation receipt, and promotion. Each part of the paper focuses on one of these stages and establishes new facts about their global evolution. Our analysis draws on the largest global database of university academics ever assembled. We hand-collect these data from historical records and modern university websites. The database contains more than half a million observations from 7,477 universities in 151 countries, spanning all disciplines and six cross-sections (1900, 1914, 1925, 1938, 1956, and 1969).¹ We further extend the data by adding information for five major scientific disciplines in the year 2000, covering prestigious universities with continuous records throughout the century. This novel dataset enables us, for the first time, to document the global long-run dynamics of gender gaps in academia.

¹For comparison, the US News ranking is based on 1,748 universities (see [here](#), accessed on July 6, 2021). The Shanghai Ranking includes 2,417 universities (see [here](#), accessed on July 6, 2021).

The distinctive strength of our data lies in extensive manual enhancements to the faculty rosters. First, we identify the gender of academics through a multi-step procedure. Second, we reconstruct individual career trajectories using a cascading algorithm that links academics across the seven cohorts. For example, we trace Margarete Bieber’s career from the University of Gießen, Germany (1925 cohort) to Columbia University, USA (1938 and 1956 cohorts).² Third, we manually recode more than 100,000 field specializations into 36 standardized disciplines and harmonize academic ranks across countries to enable the analysis of promotions. Finally, we complement the faculty rosters with publication and citation data from Clarivate Web of Science and Microsoft Academic Graph to examine gender gaps in research output and citation impact. As the data are based on complete faculty rosters, which we link to publication and citation records, we can analyze the entire population of academics who could, in principle, publish, be cited, and be promoted to full professor. This comprehensive coverage helps to overcome major selection biases that affect studies relying solely on publication and citation databases or on surveys of academics.³

In the first part of the paper, we analyze gender gaps in academic hiring. For many disciplines, countries, and universities, our data capture the first women to enter the academic profession, e.g., Katherine Coman, the first female full professor of economics in the United States (Vaughn 2004). Across all disciplines, countries, and universities, our newly collected data show that in 1900, only 227 women held academic positions, corresponding to about 1 percent of all faculty. Female representation rose slowly over the following decades: to 2 percent in 1914, 3 percent in 1925, 7 percent in 1938, 11 percent in 1956, and 11 percent in 1969. In the sciences (mathematics, physics, chemistry, biochemistry, and biology), the share of women increased from about 1 percent to 7 percent between 1900 and 1969.

We also examine hiring gaps at prestigious universities. Studying access to these institutions is particularly important, as they account for a disproportionately large share of major scientific advances. At prestigious universities, female representation in the sciences was roughly 25-50 percent lower than the average across all universities until 1969. Between 1969 and 2000, however, female shares in the sciences at these universities rose substantially—from about 4.4% to 18.7%. Despite these gains, women continue to be significantly underrepresented in the sciences at prestigious institutions. Our data also indicate that throughout the 20th century, female participation in academia was consistently lower than in the broader workforce across most countries.

We further examine how gender gaps in hiring differed across academic ranks and find particularly large disparities at the level of full professors. By 1900, universities worldwide had appointed only 114 women as full professors, corresponding to roughly 1 percent

²Margarete Bieber, an archaeologist and art historian, was the second woman in Germany to become a full professor. Because of her Jewish background, she was dismissed by the Nazi government and emigrated to the United States (Becker et al. 2024).

³For example, Huang et al. (2020) study gender inequality in science using publication data.

of all professorships. Although the share of female full professors rose over the following decades, it remained consistently lower than the overall share of women in academia. The slower rise in women’s representation at senior ranks may reflect either compositional changes over time or lower promotion prospects for women. We systematically explore gender gaps in promotions to full professor in the final part of the paper.

The global scope of our data also reveals large differences in women’s representation across countries and disciplines. Before World War I, U.S. universities employed more female academics than universities in any other country, a lead that persisted until 1969. Among prestigious universities, however, the U.S. lead in the sciences had faded by 2000, as other countries hired more women. The United Kingdom also lost its early lead, while Scandinavia, Germany, France, and Italy made substantial gains from initially low levels. Japan stands out as an exception, with little progress over the century. These patterns contrast with trends in female labor force participation in the broader workforce (e.g., Olivetti and Petrongolo 2016), suggesting that women’s academic careers were shaped by distinct institutional or cultural forces. Differences across disciplines were equally pronounced. Between 1900 and 1969, no discipline had a female share above 35%, and in most disciplines women accounted for less than 10% of academic staff. Female representation was particularly low in physics, law, veterinary medicine, architecture, and theology, but higher in pedagogy, communication studies, languages, and sports sciences. More broadly, gender gaps were substantially larger in STEM than in the humanities or social sciences.

In the second part of the paper, we investigate gender gaps in publications. A key advantage of studying academics is the availability of individual-level output measures that are comparable across time and space. Publications serve as a central metric to evaluate academics. Nonetheless, publications only partially measure true research ability, as they are influenced by preferences, discrimination, and other biases.

We measure publications over a \pm five-year interval around each cohort (e.g., 1995–2005 for scientists observed in 2000). Across all universities, between 1900 and 1969, female scientists published on average one to two fewer papers than their male colleagues—equivalent to roughly 0.2 standard deviations. Gender gaps in publications are similar if we compare men and women within the same university and cohort (e.g., Harvard in 2000) or even within the same department and cohort (e.g., physics at Harvard in 2000). At prestigious universities, the gender gap in publications is roughly 50% larger. Over the 20th century, the gap followed a U-shaped trajectory. Around 1900, women published about 0.3 standard deviations fewer papers than men. The gap widened over the first half of the century, peaking at approximately 0.45 standard deviations by the mid-1950s, before narrowing again to about 0.2 standard deviations by 2000.

To examine whether gender gaps in publications are related to gender gaps in hiring, we estimate regressions that allow the gender gap in publications to vary with the share of women in academia across countries and periods. The results reveal a U-shaped

relationship between the gender gap in publications and the share of female scientists, a pattern we term the “gender U.” In country-cohort pairs with very low female representation, the estimated gender gap is small or even positive. We refer to these as the “Marie Curie periods,” when only exceptionally talented women, such as Marie Curie, were able to obtain academic positions. As the share of female scientists increases, the gender gap widens and turns negative as academia opens to a broader pool of women (selection effect). Yet, once female representation reaches higher levels, the negative publication gap narrows again, suggesting that greater representation is accompanied by improved publishing opportunities for women and a more level playing field (empowerment effect).

In the third part of the paper, we explore whether papers published by women received fewer citations than those published by men. Citation gaps may arise not only from differential recognition of similar work but also from gender differences in research topics. To disentangle these channels, we develop a novel machine-learning approach that predicts expected citation counts based on the words in paper titles. The method estimates a regularized regression model that captures systematic variation in citation patterns across topics. We implement two versions of the model. In the first, the training sample includes all papers in the dataset. In the second, the model is trained exclusively on papers written by men, allowing us to predict how many citations each paper would have received had it been authored by a man. The model trained on all papers may provide more accurate predictions of realized citation counts, whereas the model trained only on male-authored papers addresses the concern that fields with substantial female contributions may be systematically undercited relative to their underlying quality, precisely because many of the papers in those areas were written by women.

Before World War I, papers by female authors received around 0.2 standard deviations fewer citations than those by male authors. During the interwar and post-war periods, this gap narrowed to roughly 0.1-0.15 standard deviations. By 2000, it had declined further to around 0.05 standard deviations. The estimated gender gaps in citations remain similar if we control for our novel predicted citation measure and other paper characteristics. This indicates that gender gaps in citations did not stem from gender differences in the number of coauthors, journal placement, or, most importantly, citation potential. Instead, women’s papers received fewer citations due to other biases.

In the fourth part of the paper, we investigate how women advanced in their academic careers by studying gender gaps in promotions. Until the late 1960s, women were about 10-20 percentage points less likely than men to be promoted to full professor. By 2000, the gender gap in promotions in the sciences narrowed to about 7 percentage points. Gender gaps in promotions are very similar if we compare men and women who entered the data in the same university and cohort (e.g., Berkeley in 1900) or even the same department and cohort.

A potential explanation for women’s lower promotion prospects could be the gender gaps in publications and citations that we uncover in the second and third parts of the

paper. However, even when we control for scientists’ publication and citation records, we estimate very similar gender gaps in promotions. Strikingly, this unexplained gender gap in promotions is larger than the effect of a publication record three standard deviations worse. Overall, these patterns suggest that the slower career progression of women over much of the 20th century did not stem from sorting into institutions or departments with poorer promotion prospects, or from differences in publication or citation records. Instead, the gender gap in promotions was likely driven by other biases.

This paper contributes to the literature on the allocation of talent in innovation and the production of ideas. Existing work shows that children from poorer backgrounds are less likely to become inventors (Bell et al., 2019) or scientists (Abramitzky et al., 2024), and that researchers’ identities shape the questions they study (Einiö et al., 2019; Abramitzky et al., 2024; Dossi, 2024). We extend this literature by documenting the global misallocation of female talent in academia over the 20th century. Our evidence identifies a large group of “lost Marie Curies:” women whose entry into academic careers was delayed or prevented, and whose scientific contributions were under-cited.

Our paper also contributes to the literature on gender gaps in academia. In economics, female-authored papers receive more citations than male-authored papers, consistent with higher publication thresholds for women (Card et al., 2022), and women were less likely to be elected as fellows of major academies until the late 1970s (Card et al., 2023). Women receive less credit for collaborative work (Sarsons, 2017b; Sarsons et al., 2021), are treated differently in seminars (Dupas et al., 2026), are more often omitted from reference lists (Koffi, 2025), are held to higher writing standards (Hengel, 2022), and spend more time on tasks with low promotability (Babcock et al., 2017). The gender composition of evaluation committees affects women’s promotion prospects (Bagues et al., 2017), and physicians judge female surgeons more harshly after adverse outcomes (Sarsons, 2017a). Beyond economics, women are less likely to receive authorship credit (Ross et al., 2022), mothers of young children experienced lower productivity in mid-20th-century U.S. science (Moser and Kim, 2021), and a higher share of female undergraduates shifts professors’ research toward gender-related topics (Truffa and Wong, 2022).

We advance this literature in three ways. First, we provide comprehensive global evidence for more than 150 countries over the entire 20th century. Second, we study the full population of academics rather than selected samples of publishing scientists or famous scholars. This is crucial because women were less likely to publish and receive recognition in the first place. Third, we analyze hiring, publications, citations, and promotions within a unified dataset. This allows us to study how gender gaps interact across career stages, including how women’s entry into academia shaped subsequent publication gaps. This finding echoes evidence that positive selection of women affects gender gaps in wages, both in a large multinational firm (Ashraf et al., 2022) and in the broader U.S. economy (Mulligan and Rubinstein, 2008; Hsieh et al., 2019).

Our work also relates to the literature on gender gaps in specific high-skilled professions, including MBA graduates (Bertrand et al., 2010), executives (Bertrand and Hallock, 2001; Gayle et al., 2012; Albanesi et al., 2015), lawyers (Azmat and Ferrer, 2017), pharmacists (Goldin and Katz, 2016), and engineers (Roussille, 2021).⁴ Studying academia allows us to extend this lens globally and over the long run, because faculty rosters for much of the 20th century are systematically available worldwide, and individual-level output measures are observable and comparable across time, country, and discipline.

Finally, we contribute to the literature that uses machine-learning and text-based methods in economics (e.g., Hoberg and Phillips, 2016; Hansen et al., 2018; Iaria et al., 2018; Gentzkow et al., 2019).⁵ In particular, we contribute to research that uses these methods to detect gender gaps. Closest to our approach, Koffi (2025) uses text similarity to estimate gender gaps in citations. This alternative approach relies on text similarity to identify which papers should have been cited. In contrast, our new method uses a regularized regression to predict each paper’s citation potential, which allows us to estimate overall citation gaps, rather than gaps among the most similar papers. A further advantage of our method is its suitability in settings where female-authored papers are rare and therefore unlikely to appear among the closest matches in similarity-based comparisons. It is also computationally feasible in large samples.⁶ More broadly, our new method can be applied in any setting where unstructured text, such as paper titles or performance reviews, can be used as a proxy for the quality of work being evaluated.

1 A New Database of University Academics

At the core of this paper is the largest database of university academics ever assembled. We hand-collect faculty rosters for a very large number of universities worldwide from the historical publication *Minerva Jahrbuch der Gelehrten Welt* and from modern university websites. We then link these data to detailed publication and citation records from Clarivate Web of Science and Microsoft Academic Graph. Throughout the paper, we present results for three samples:

- Sample 1: all universities, all disciplines, 1900-1969.
- Sample 2: all universities, sciences (mathematics, physics, chemistry, biochemistry, and biology), 1900-1969, with publication and citation data.
- Sample 3: prestigious universities, sciences (mathematics, physics, chemistry, biochemistry, and biology), 1900-2000, with publication and citation data.

⁴Related research documents gender gaps in the broader labor market but has largely focused on individual countries and short time horizons (for surveys, see Altonji and Blank, 1999; Bertrand, 2011; Blau and Kahn, 2017; Bertrand and Duflo, 2017). Goldin’s work on the United States is a notable exception in historical scope (e.g., Goldin, 1989, 1990).

⁵See Ash and Hansen (2023) for a survey.

⁶The alternative approach by Koffi (2025) requires calculating pairwise similarities between all papers, implying that the computational burden grows quadratically with the number of papers.

Sample 1 enables us to study gender gaps in hiring and promotion across all disciplines and universities worldwide. Sample 2 focuses on the sciences and additionally enables us to examine gender gaps in publications and citations, since the scientific disciplines had already established an international culture of publishing and citing research comparable to modern standards by 1900. Finally, Sample 3 extends the analysis to 2000 for prestigious universities. Examining access to these institutions is particularly important, as they account for a disproportionately large share of major scientific advances.

1.1 Hand-collection of Faculty Rosters 1900-2000

Historical Faculty Rosters for the Years 1900-1969

For the period 1900 to 1969, we digitize faculty rosters from *Minerva Jahrbuch der Gelehrten Welt* (hereafter *Minerva*). Prior to the advent of the internet, *Minerva* represented the most comprehensive global directory of university academics. Its publishers systematically collected information from ministries of education, university administrators, and individual scholars to ensure near-complete coverage of the academic population worldwide.⁷ *Minerva* was published in volumes that provide worldwide cross-sections of university academics. We digitize six volumes covering the years 1900, 1914, 1925, 1938, 1952/56, and 1966/69 (see Figure 2 for a sample page). These cohorts were chosen based on data availability and to align with major historical turning points in 20th-century academia: 1900 marks the beginning of the Web of Science data, which enables an analysis of publication and citation gaps; 1914 captures the cohort just before World War I (WWI); 1925 represents the period at the end of the academic boycott against scholars from the Central Powers (Iaria et al. 2018); 1938 precedes World War II (WWII); and 1956 and 1969 represent the only postwar cohorts ever produced by *Minerva*.⁸ For the remainder of the paper, we refer to these years as cohorts.

To digitize the *Minerva* data, we scan approximately 10,000 pages of the relevant volumes and process them using optical character recognition (OCR) software. We then parse the largely unstructured OCR output to extract structured information such as names, affiliations, academic specializations, and academic titles. We then hand-check each record to correct OCR errors and verify accuracy. This procedure yields around half a million person-cohort observations (Table 1), covering 7,477 universities in 151 countries.

⁷A Nature article compared the French *Annuaire Général des Universités* to *Minerva* and noted that the former “is akin to the well-known *Minerva*.... It is, however, very much less exhaustive” (Nature 1930). To our knowledge, no other source offers comparable worldwide coverage of academics over such an extended period. We benchmark *Minerva* in two ways. First, we show that the number of universities included in *Minerva* closely matches the number of universities listed in the World Higher Education Database (WHED), which has been designed as a census of worldwide universities (WHED 2024; see Appendix Table A.3). Because the WHED lacks microdata on individual academics, we complement this comparison with additional benchmarking exercises using smaller datasets that include individual-level information for selected universities and periods. These exercises indicate that *Minerva* captured a large share of the global academic population (Appendix A.6.).

⁸Because the number of academics expanded substantially over time, *Minerva* published each of the postwar cohorts in two installments. Throughout the paper, we refer to these cohorts by the later publication year, e.g., as 1956 for the 1952/56 volume.

Figure 1 displays the global distribution of academics across cities, while Appendix Figure B.1 shows the evolution of the share of academics by continent over time.⁹

The data cover the full spectrum of higher education institutions: traditional universities such as Harvard or the University of Tokyo, technical universities such as MIT or École Polytechnique, mining universities such as Freiberg Mining Academy, and theological universities such as Pontificia Università Gregoriana in Rome. Virtually all Ph.D.-granting institutions are included in the data. For example, the data contain academics in 1,540 universities in the United States, 309 in the United Kingdom, 281 in Germany, and 351 in France (Appendix Table A.3 reports the number of universities by country). For most universities, the data include all assistant, associate, and full professors, as well as honorary professors and, in some cases, individuals in research or teaching roles.¹⁰

Modern Faculty Rosters For The Year 2000

For the year 2000, we digitize faculty rosters from archived university websites using the Internet Archive’s Wayback Machine. We focus on five scientific disciplines—mathematics, physics, chemistry, biochemistry, and biology—and collect data for 249 prestigious universities across 34 countries. These universities either appear in all six historical *Minerva* cohorts or rank among the top 100 institutions in the 2020 Shanghai Ranking, allowing us to extend the historical *Minerva* data into the modern period. Between 1900 and 1969, roughly 51% of all scientists worldwide were affiliated with these 249 institutions, and they accounted for 68% of all publications. These universities therefore represent a substantial share of global scientific activity during this era.¹¹

Enhancements of Faculty Roster Data

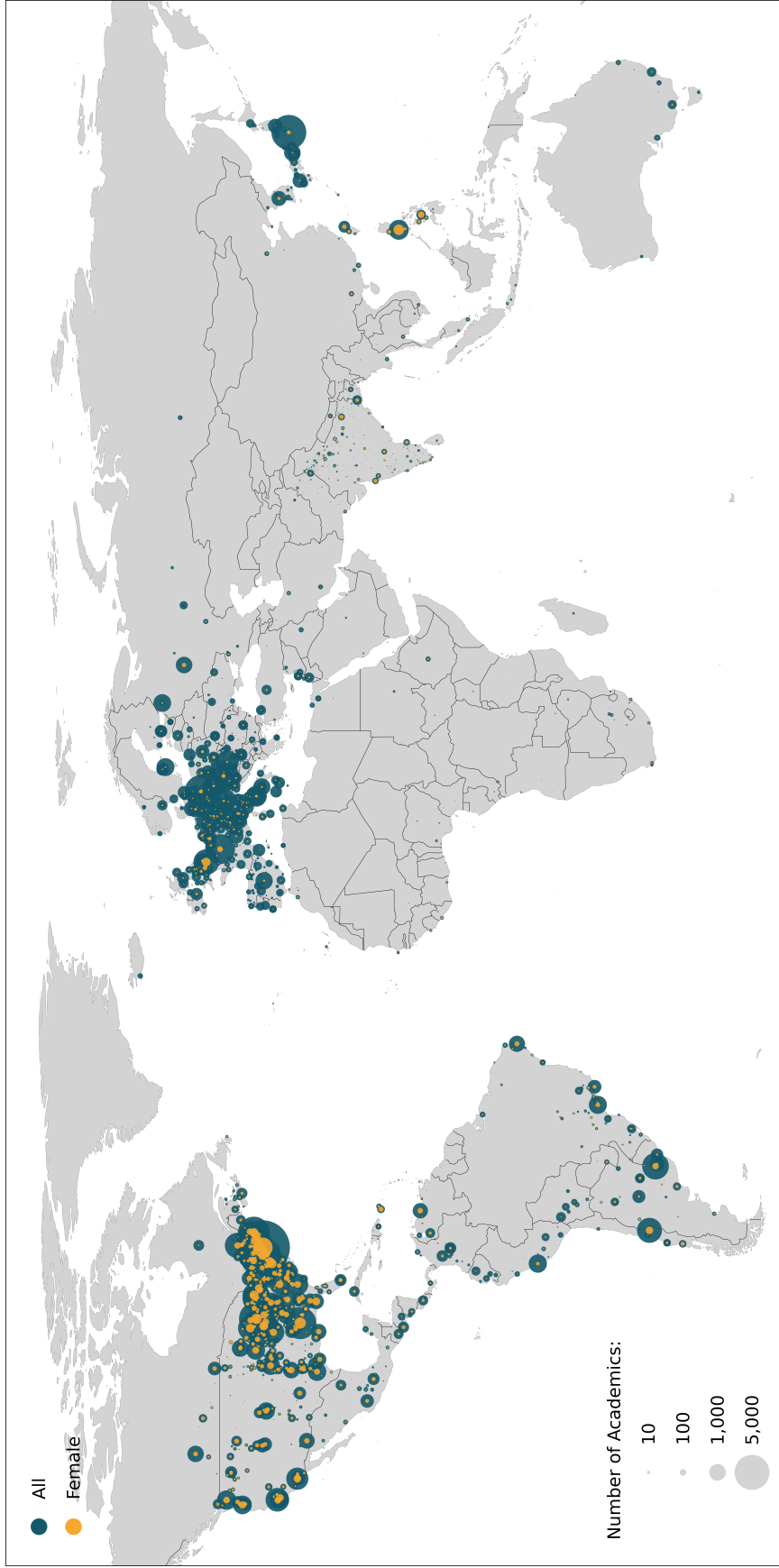
We make a large number of manual enhancements to the faculty rosters. First, we recode thousands of academic ranks (e.g., “professor,” “chargé de cours,” or “incaricato”) into eleven harmonized categories (e.g., assistant professor, full professor, or teaching position; see Appendix A.1.1.). Second, we manually recode more than 100,000 field labels (e.g., “quantum theory” or “physique des particules élémentaires”) into 36 disciplines (e.g., physics; see Appendix A.1.2.). Third, we consolidate multiple appointments held by the same academic within a university or city into a single observation (see Appendix A.1.3.).

⁹Relative to existing datasets used in economics, our data cover a substantially larger number of academics and universities. For example, the notable data collection effort by De la Croix et al. (2023) includes 47,897 academics across 198 universities for the period 1000 to 1800.

¹⁰Coverage varies slightly across institutions. For some smaller universities, particularly in India, *Minerva* reports only the total number of professors without listing individual names. In other cases, the source lists named professors but reports other teaching positions only in aggregate form (e.g., “10 lecturers”). Across all cohorts, *Minerva* records 498,525 faculty members with identifiable names (Table 1) and an additional 108,398 faculty positions for which names are not reported.

¹¹The sample for the year 2000 includes 69 universities in the United States, 31 in Germany, 24 in the United Kingdom, 22 in Italy, 21 in France, and 9 in Switzerland. It further comprises 7 universities each in Australia, Austria, and Canada; 5 each in Belgium and the Netherlands; 4 each in Denmark and Sweden; and 3 each in Japan and Ireland. In addition, the sample includes 2 universities in Argentina, Finland, Hungary, New Zealand, Portugal, and Spain, and 1 university in each of the following countries: Bulgaria, Chile, Croatia, Czech Republic, Greece, India, Israel, Norway, Pakistan, Peru, Poland, Romania, Russia, Serbia, Singapore, and Uruguay.

Figure 1: Academics 1900-1969



Notes: The map shows the total number of person-cohort observations in the 1900-1969 cohorts by city. All academics are represented by blue dots, female academics are represented by orange dots. The data were collected by the authors, see Section 1 for details.

Table 1: Summary Statistics

	All Academics	All Gender Coded	Female	Male
<i>Sample 1: All Universities, all disciplines, 1900-1969</i>				
Number of academic - cohort observations	498,525	411,302	35,441	375,861
Number of universities	7,477	5,503	2,399	5,136
Number of departments	37,083	30,677	8,723	29,117
Female %		8.62	100.00	0.00
<i>Sample 2: All Universities, sciences, 1900-1969</i>				
Number of academic - cohort observations		67,618	3,714	63,904
Number of universities		2,119	880	2,050
Number of departments		6,429	1,648	6,156
Female %		5.49	100.00	0.00
Publications		4.01	1.65	4.15
<i>Sample 3: Prestigious Universities, sciences, 1900-2000</i>				
Number of academic - cohort observations		88,537	11,378	77,159
Number of universities		249	248	249
Number of departments		1,202	1,009	1,200
Female %		12.85	100.00	0.00
Publications		9.84	6.37	10.35

Notes: The Table shows summary statistics at the academic-cohort level. Sample 1 includes academics in all universities and disciplines from 1900 until 1969. Sample 2 includes academics in all universities in mathematics, physics, chemistry, biochemistry, and biology from 1900 until 1969. Sample 3 includes academics in prestigious universities in mathematics, physics, chemistry, biochemistry, and biology from 1900 until 2000. The data were collected by the authors from *Minerva*, university websites, *Clarivate Web of Science*, and *Microsoft Academic Graph*; see Section 1 for details.

Fourth, we link academics across cohorts using a cascading matching algorithm (see Appendix A.1.4.). Fifth, for entries listing only surname and initials, we conduct manual web searches to identify full first names (see Appendix A.1.5.).¹² Sixth, we construct consistent university identifiers by linking universities across cohorts and tracking university mergers and splits throughout the 20th century.

Identifying the Gender of Academics

We develop a five-step procedure to identify the gender of academics on a global scale. In the first step, we draw on explicit gender markers in the *Minerva* directories, such as formal titles (e.g., names preceded by Mlle., Lady, Lord, Cardinal), or from the department websites, including photographs and pronouns in biographical descriptions. In all subsequent steps, we rely on first names to identify the gender of academics.

In the second step, we classify more than 100,000 ‘first name’-country pairs with gender-api.com. Gender-api.com distinguishes the gender associated with first names at the country level (e.g., Andrea is typically male in Italy but female in many other countries), and assigns a gender probability to each ‘first name’-country pair.¹³

¹²All results are robust to excluding this step.

¹³At the time of writing, Gender-api.com was the best-performing name-to-gender inference service (Santamaría and Mihaljević 2018).

In the third step, two research assistants (one male and one female) independently classify ‘first name’-country pairs that gender-api.com classified as less than 100% male. The research assistants are instructed to classify only cases for which they can assign gender with certainty. If the two assistants’ classifications coincide, the procedure ends.

In the fourth step, we process the remaining cases that gender-api.com classified as less than 100% male by searching the ‘first name’-country pairs using a Google image search. A research assistant then classifies each ‘first name’-country pair as male or female depending on whether the image search predominantly displays men or women. For example, gender-api and the research assistants could not identify the gender of “Hadmar” in Austria. A Google search for “Hadmar Austria” returned only images of men (Appendix Figure A.1). We therefore code Austrian academics named “Hadmar” as male.

In the fifth step, using targeted Google searches, we hand-check individual academics who appear misclassified (see Appendix A.2.2.). Misclassifications arise primarily because the predominant gender associated with some names changes over time. For example, the French name “Camille” is historically male but becomes increasingly female in later cohorts.¹⁴ While the manual steps significantly increase data quality, none of the results change without steps 3-5. Throughout the paper, analyses are based on the sample of academics for whom gender is successfully determined (see Table 1). Overall, we assign gender to 82.5% of academics. The remaining unidentified cases arise predominantly in universities that report faculty only by initials, which prevents reliable gender attribution.

Examples of Academics in the Database

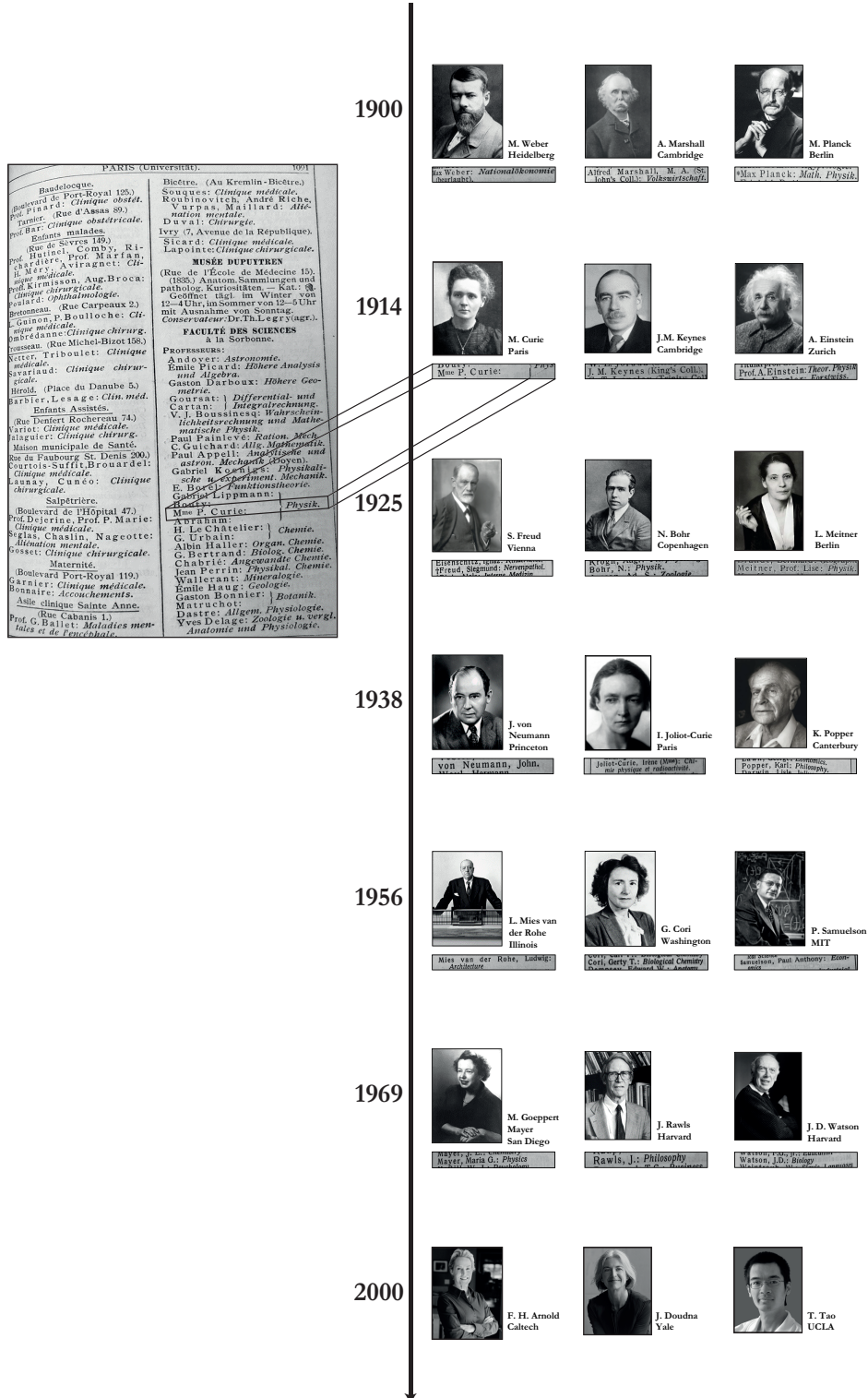
Figure 2 shows three exemplary academics for each cohort. The selection showcases some of the data’s country, discipline, cohort, and gender dimensions. However, it does not do justice to the tens of thousands of academics who have contributed to the progress of knowledge. For 1900, the data include the economist Alfred Marshall (University of Cambridge), the physicist and Nobel laureate Max Planck (University of Berlin), and the sociologist Max Weber (University of Heidelberg).

Examples for 1914 are John Maynard Keynes (University of Cambridge), Albert Einstein (ETH Zürich), and arguably the most famous woman in our data, Marie Curie (Université de Paris). Together with her husband, she conducted pioneering research on radioactivity and was the first woman to win the physics Nobel Prize in 1903. Despite this, she was not awarded a professorship at the Sorbonne. Only after her husband had tragically died, she finally became the first female full professor at the Sorbonne, five years after winning her first Nobel Prize, and two years before she won her second for her contributions to chemistry (McGrayne 1998).¹⁵

¹⁴In this step, the Google search uses full identifying information: surname, first name, and discipline or university, rather than only a first-name-country pair.

¹⁵Strikingly, Curie is listed in *Minerva* as Mme P[ierre] Curie at a time when she had won two Nobel Prizes (Figure 2). Throughout her career, she faced obstacles because of her gender. In 1911, her nomination to the Academy of Sciences was met with resistance, epitomized by physicist Émile Amagat’s

Figure 2: Examples of Academics in Database



Notes: The Figure shows three examples of notable academics for each of the seven cohorts of academics.

In 1925, the data list the physics Nobel laureate Niels Bohr (University of Copenhagen), the founder of psychoanalysis Sigmund Freud (University of Vienna), and Lise Meitner (University of Berlin). Meitner was the second woman to earn a physics Ph.D. at the University of Vienna. During her Post-Doc at the University of Berlin, she was unpaid and had to run her experiments in a converted carpenter's shop in the cellar because—as a woman—she was barred from entering the main laboratory. Together with Otto Hahn, she discovered nuclear fission. The Nobel laureate Wolfgang Pauli commented that “Hahn and Meitner were great friends, but when they talked, she was superior.” In 1945, the Nobel Prize was awarded to Hahn alone, neglecting Meitner's role, a decision contemporaries called a “stupidity of the Swedish Academy” (Kricheldorf 2014, p. 219).

Examples for 1938 are the mathematician John von Neumann (IAS Princeton), the philosopher Karl Popper (University College Canterbury, NZ), and Irène Joliot-Curie. She was only the second woman to win a Nobel Prize in chemistry, more than 20 years after her mother. After winning the Nobel Prize, her fellow Nobel laureate and husband Frédéric Joliot-Curie was admitted to the French Academy of Sciences, while she was rejected every time she applied (McGrayne 1998, p. 140).

Examples for 1956 are Ludwig Mies van der Rohe (Illinois Institute of Technology), one of the pioneers of modernist architecture, the economist Paul Samuelson (MIT), and Gerty Cori (Washington University). Cori was the first woman to win the physiology/medicine Nobel Prize in 1947 (and the third woman to win a science Nobel Prize). Despite her talent, Cornell, Toronto, and Rochester refused to hire her while offering professorships to her husband and fellow Nobel laureate Carl Cori. In 1931, Washington University made them a joint offer, but Gerty was hired as a research associate while Carl was hired as a full professor (Shepley 2008, McGrayne 1998, pp. 102).

The 1969 cohort includes the moral philosopher John Rawls (Harvard), the biologist and discoverer of the double-helix structure of DNA James Watson (Harvard), and the theoretical physicist Maria Goeppert Mayer (UC San Diego), who proposed the nuclear shell model of the atomic nucleus. “[S]he worked for thirty years ... for three American universities ... as an unpaid volunteer” (McGrayne 1998, p. 175). Johns Hopkins and Columbia refused to hire her because of nepotism restrictions (her husband was a chemist). Only in 1960, at the age of 54, and ten years after completing her most important research, was she appointed full professor at UC San Diego (Wigner 1972). In 1963, she became the second woman to win the physics Nobel Prize, 60 years after Marie Curie. The 2000 cohort includes the chemistry Nobel laureate Frances H. Arnold (Caltech) and the biochemist Jennifer Doudna (Yale), who co-discovered a method of gene editing using CRISPR/Cas9. Together with Emmanuelle Charpentier, she was awarded the Nobel Prize in 2020, the first all-female winners of the chemistry Nobel Prize. Another example from the 2000 cohort is the mathematician Terence Tao (UCLA), who won the Fields Medal in 2006.

argument “Women cannot be part of the Institute of France.” Despite the efforts of some of France's greatest scientists, she narrowly lost the membership election to a male competitor (Curie 1938, pp. 277).

1.2 Publication and Citation Data

To study gender gaps in publications and citations, we augment the faculty rosters with publication and citation data from Clarivate Web of Science. For any result based on publications and citations, we focus on five academic disciplines: mathematics, physics, chemistry, biochemistry, and biology. There are three reasons for this. First, these disciplines have particularly good coverage in the Web of Science. Second, they had already established the culture of publishing in scientific journals by 1900, and the publishing process was similar to today’s. Third, the publishing process was international (Iaria et al. 2018). For the years of our study, the Web of Science contains papers in 14,191 journals in these disciplines. Naturally, the coverage of the Web of Science is not uniform across countries and disciplines, or over time. This does not affect our estimates as we control for cohort-discipline-country (or finer) fixed effects in all regressions.

We link academics with their publications using a cascading matching algorithm (see Appendix A.4.). The matches are based on the academic’s surname, first name or initials (depending on whether first names are available), country, city, and discipline.¹⁶ To harmonize affiliations across the faculty rosters and the Web of Science, we rely on Google Maps API (see Appendix A.3.2.) to extract cities and countries for each of the hundreds of thousands of unstructured affiliations. For example, we extract the city “Cambridge” and the country “UK” from the affiliation “Cavendish Lab., Cambridge University, UK.”

The matching always relies on the primary discipline of an academic (e.g., biology) to reduce false positives. As the Web of Science assigns disciplines (e.g., biology or general science) only at the journal level, we develop a machine-learning classifier to assign disciplines to individual papers (see Appendix A.3.3.). The classifier, an L2-regularized multinomial logit model, predicts a discipline for each paper based on the unigrams, bigrams, and trigrams from the titles of the 59% papers published in journals assigned to only one discipline.

We consider publications in a \pm five-year window around the year of the corresponding cohort. For example, for scientists in the 2000 cohort, we match papers published between 1995 and 2005.¹⁷ In the rare cases where two or more scientists have identical names and work in the same discipline, we assign the paper proportionally to each scientist.¹⁸

¹⁶For many papers, the Web of Science only reports the initials of authors. In addition, for some papers, the Web of Science does not report affiliations, even though the original paper actually lists an affiliation. In some of these cases, Microsoft Academic Graph (MAG) contains the relevant affiliation. We therefore enrich the affiliation information with data from MAG (see Appendix A.3.2.).

¹⁷We use a \pm five-year window because scientists do not necessarily publish every year. Concerns that the matches of female academics may be affected by surname changes from marriage are mitigated by various factors. First, the faculty rosters list academics who are at least assistant professors. Hence, most married women were already married when appearing on the rosters. Second, marriage rates for female academics in the early part of the 20th century were low, e.g., 18% in 1921 and 26% in 1938 for scientists in the United States (Rossiter 1982, p. 140). Furthermore, estimated gender gaps remain unchanged if we match publications in a \pm three-year window, which reduces the probability of name changes.

¹⁸Results are robust in a sample of scientists who were unique in terms of last name, first initial, and discipline in any university of the world (Table 3).

2 Gender Gaps in Hiring

Hiring of Women Over Time

In the first part of the analysis, we present the first-ever global evidence on the long-run evolution of gender gaps in hiring at universities. We show results for the following samples: Sample 1: all universities, all disciplines, 1900-1969; Sample 2: all universities, sciences (mathematics, physics, chemistry, biochemistry, biology), 1900-1969; and Sample 3: prestigious universities, sciences, 1900-2000. We report the absolute number of male and female academics (left-hand panels of Figure 3), as well as female shares among all academics and among full professors (right-hand panels of Figure 3).

Across all countries and disciplines (Sample 1), our newly collected data show that by 1900 only 227 women had been hired, a share of about 1% (Figure 3, panel (a)). In the following decades, the share of women increased, in particular between 1925 and 1938, i.e., before WWII. By 1969, a total of 17,204 women had been hired across all universities and disciplines, a share of about 11%, still nowhere close to equal representation.

We also explore changes in female shares among full professors. All over the world, full professor is the highest academic rank and carries unique privileges, including particularly high job security and salaries. In addition, full professor is the most comparable academic rank across different university systems. In 1900, only 114 women worked as full professors across all universities, a share of about 1%. In the following decades, the share of women among full professors increased, and by 1969 reached about 8%.

The slower increase in the share of women among full professors compared to all academics may indicate that women were less likely to be promoted, but may also reflect compositional changes. For example, if a higher share of women was hired in later cohorts, it could take time for these women to rise through the ranks. We systematically explore gender gaps in promotions to full professor in Section 5.

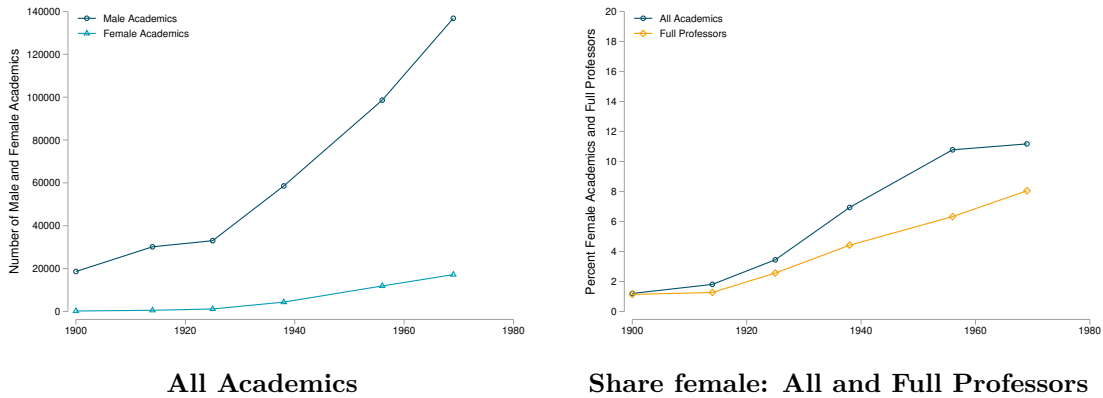
In the sciences (mathematics, physics, chemistry, biochemistry, and biology), female shares were substantially lower. Across all universities, the share increased from about 1.1% to about 6.6% between 1900 and 1969 (Figure 3, panel (b)). In prestigious universities, female shares in the sciences were even lower until 1969. This suggests that women faced particularly large hurdles to obtaining positions in prestigious universities. Between 1969 and 2000, female shares in the sciences in prestigious universities increased substantially: from about 4.4% to about 18.7% and from 3.3% to 8.8% among full professors (Figure 3, panel (c)). Despite this large increase, women were still heavily underrepresented in prestigious universities in 2000, especially among full professors.

Hiring Gaps Across Countries

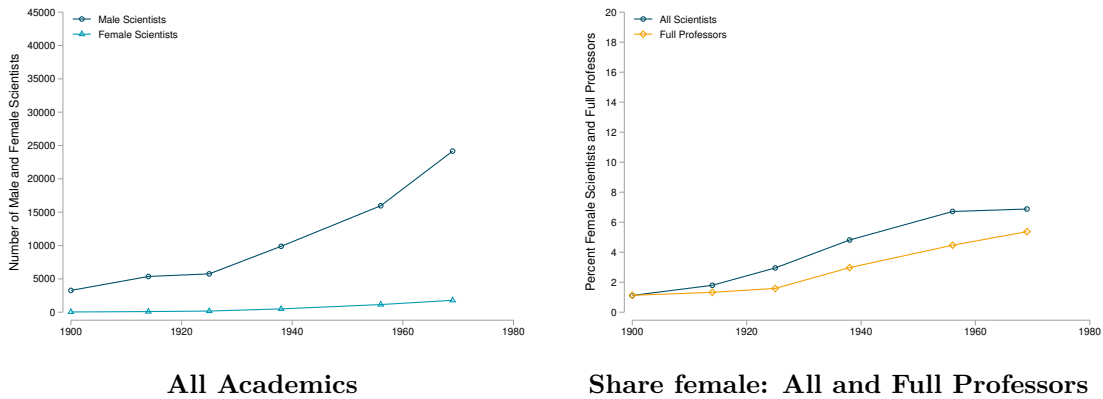
The aggregate statistics hide significant heterogeneity between geographical regions (Figure 4). Until 1925, North America had much higher female shares than other regions (panel (a)). By 1938, the female share in the few African universities had caught up. Asia and South America had very low female shares until WWII, but then their shares

Figure 3: Absolute and Relative Number of Female Academics over Time

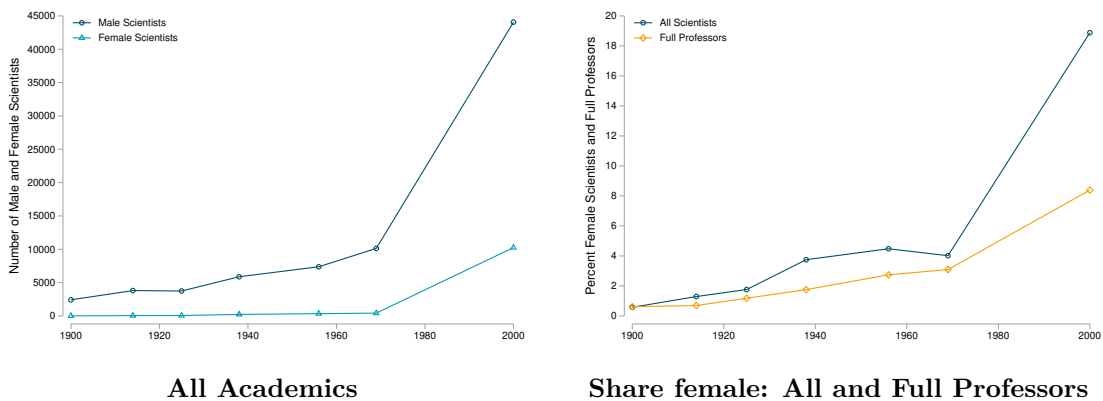
(a) Sample 1: All universities, all disciplines, 1900-1969



(b) Sample 2: All universities, sciences, 1900-1969



(c) Sample 3: Prestigious universities, sciences, 1900-2000



Notes: The Figure shows the absolute and relative number of female academics over time. Panel (a) shows the number of male and female academics (left sub-panel) and female shares (right sub-panel) in all universities and disciplines until 1969. Panel (b) shows the absolute number of male and female scientists (left) and female shares (right) in all universities in the sciences (mathematics, physics, chemistry, biochemistry, and biology) until 1969. Panel (c) shows the absolute number of male and female scientists (left) and female shares (right) in prestigious universities in the sciences until 2000. The data were collected by the authors from *Minerva* and department websites; see Section 1 for details.

increased to the level of North America and Africa. Europe and Oceania had very low female shares until 1969.

The variation in female shares at the country-level was even higher. We explore country-level differences by showing the evolution of female shares for a selected set of

countries, those that employed the largest number of academics throughout the 20th century. Before WWI, universities in the United States and, to a lesser extent, the United Kingdom hired more women than universities in any other country in the world. In the sample of all universities and disciplines (Sample 1), the dominant role of the United States persisted until 1969 (panel (b)).¹⁹ In the sciences (Sample 2), universities in the United States and the United Kingdom hired larger shares of women than other countries until WWII. They were then overtaken by France (panel (c)). In prestigious universities (Sample 3), only U.S. universities had hired women by 1900, while in the other seven countries, no prestigious university had hired any woman. By 1914, prestigious universities in the United Kingdom had hired a similar share of women as the United States. The early lead of the United States and the United Kingdom lasted until 1938, when other countries began to overtake them (panel (d)).

Sweden, and to a lesser extent Germany, had very low female shares until 1969 but increased their shares substantially in the three decades until 2000.²⁰ Italy and France had low shares before WWI, were ranked in the middle until 1969, but then increased their shares substantially in the three decades until 2000.²¹ Japan is a clear outlier: female shares at prestigious universities were very low during the first decades of the 20th century and, unlike in other advanced countries, increased only slightly until 2000.

No comparable data document gender gaps in hiring at the global level across the entire twentieth century, either for specific occupations or for the labor force as a whole. However, we correlate gender gaps in academic hiring to female labor force participation for selected countries and periods (Appendix Figure B.3). We find no systematic relationship with trends in female employment in the general population. This suggests that the careers of women in academia evolved differently from careers in lower-skilled professions.

Hiring Gaps Across Selected Universities

Our detailed data also allow us to explore university-level variation in hiring gaps. The presentation of a few university-level figures cannot do justice to the many excellent universities around the world (too many to be plotted in a single figure). To select universities for this exercise, we rely on the well-known Shanghai Ranking of universities (Shanghai Ranking 2020). We choose the highest-ranked universities in each country and report the average female shares from 1900 to 2000.²²

¹⁹The early US lead is partly due to women’s colleges, which hired more women. However, by 1938, other U.S. universities were also hiring a higher share of women than universities in other countries (Appendix Figure B.2).

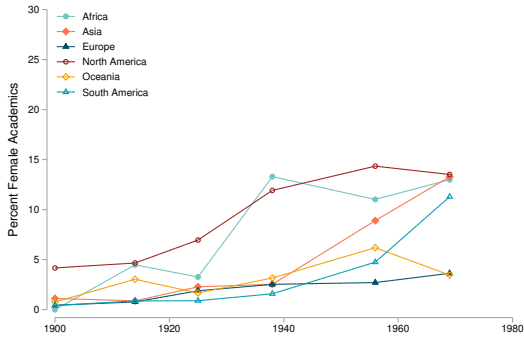
²⁰Austria and Finland show a similar development, as shown in Appendix Figure B.2.

²¹Other Romance-speaking countries had a similar development. Female shares in 2000 were 51% in Argentina, 37% in Spain, and 28% in Portugal (Appendix Figure B.2).

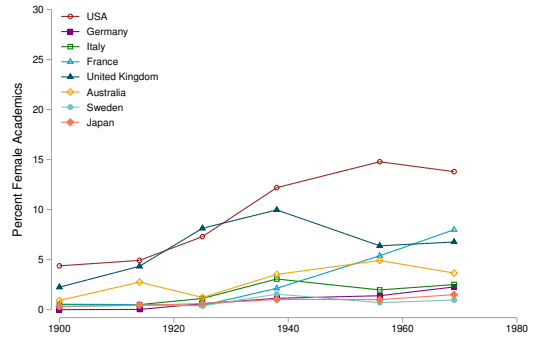
²²The Shanghai Ranking ranks universities as of 2020. In many countries, e.g., the United States, the ranking has remained stable since 1900. In other countries, the ranking has changed substantially. To accurately reflect the most important institutions during the 20th century, we deviate from the Shanghai Ranking for two countries. For Germany, we show the University of Berlin (Humboldt), the premier institution until WWII, rather than the University of Bonn. In France, several reorganizations of universities occurred during the 20th century. We thus show the Université de Paris and the Université de Grenoble.

Figure 4: Percent of Female Academics by Country over Time

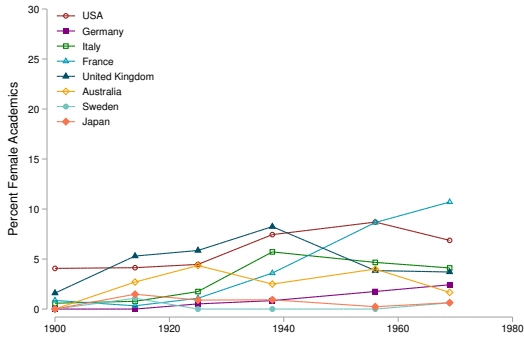
(a) *Sample 1: All unis, all disciplines, 1900-1969*



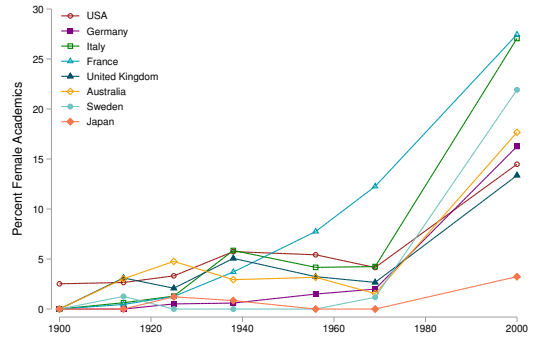
(b) *Sample 1: All unis, all disciplines, 1900-1969*



(c) *Sample 2: All unis, sciences, 1900-1969*

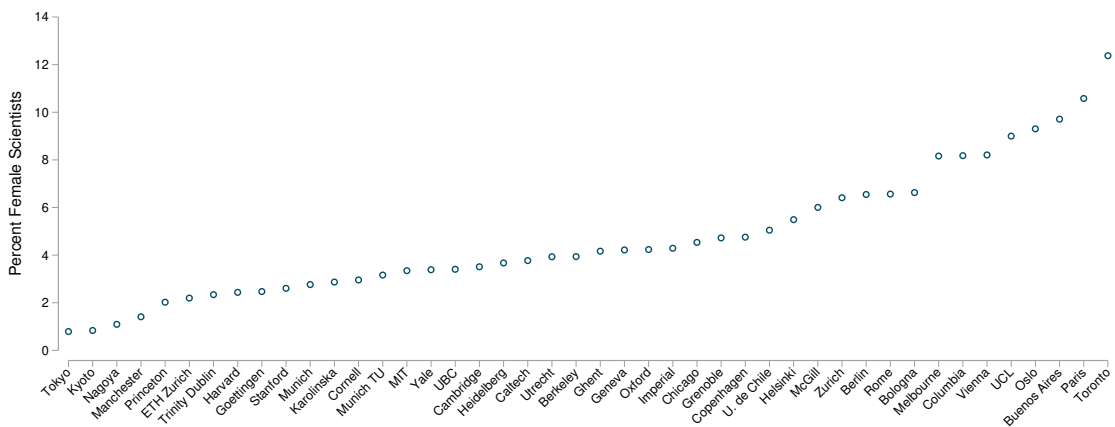


(d) *Sample 3: Prestigious unis, sciences, 1900-2000*



Notes: The Figure shows the percentage of female academics by continent and country and over time. Panel (a) shows female shares by continent in all universities and disciplines until 1969. Panel (b) shows female shares by country in all universities and disciplines until 1969. Panel (c) shows female shares by country in all universities in the sciences (mathematics, physics, chemistry, biochemistry, and biology) until 1969. Panel (d) shows female shares by country in prestigious universities in the sciences until 2000. The data were collected by the authors from *Minerva* and department websites; see Section 1 for details.

Figure 5: Percent of Female Scientists by University 1900-2000



Notes: The Figure shows the percentage of female scientists (mathematics, physics, chemistry, biochemistry, and biology) by university. Universities were selected from Sample 3 as explained in the text. We calculate percentages of female academics at the cohort and university-level, e.g., MIT in 2000, and then average the percentages over the seven cohorts (so that each cohort receives the same weight, independently of the number of academics in that cohort). The data were collected by the authors from *Minerva* and department websites; see Section 1 for details.

We report data on universities from various countries: ten universities from the United States; five each from Germany and the United Kingdom, three each from Canada, Japan, and Switzerland; two from France and Italy; and one from Argentina, Australia, Austria, Belgium, Denmark, Finland, Ireland, the Netherlands, Norway, and Sweden. The figure shows large differences in female shares across universities. Even within countries, university-level female shares vary widely. For example, over the 20th century, Columbia hired, on average, around 8% women in the sciences, while Princeton only hired around 2%.

Hiring Gaps Across Disciplines

Our data also enable us to document differences in hiring gaps across disciplines. At the aggregate level, female shares in the humanities and social sciences increased from 1-2% in 1900 to around 15% in 1969. In STEM disciplines, female shares were much lower and increased from about 1% in 1900 to only 5% in 1969 (Figure 6, panel (a)).

At a disaggregate level of disciplines, female shares varied substantially. In the first decades of the 20th century, most disciplines had very low (below 5%) female shares. For most disciplines, female shares remained below 10% until 1969, with particularly low shares in law, physics, and philosophy. However, some disciplines, e.g., pedagogy and languages, had higher female shares. By 1969, female shares in these disciplines had risen to 25% and 20%, respectively (Figure 6, panel (b)). Appendix Figure B.4 shows additional disciplines.

In the science sample (mathematics, physics, chemistry, biochemistry, and biology), female shares were lower than in many other disciplines but varied substantially across disciplines (Figure 6, panel (c)). In the sample of prestigious universities, female shares were even lower until 1969. In the last three decades of the 20th century, they increased substantially and by 2000 reached around 12% in physics, 15% in mathematics, 19% in chemistry, 26% in biochemistry, and 27% in biology (Figure 6, panel (d)).

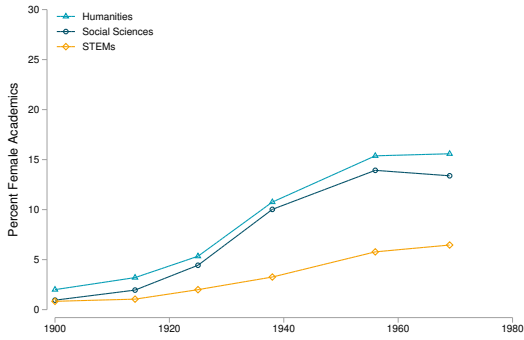
3 Gender Gaps in Publication Output

In this section, we examine gender gaps in publications. A key advantage of studying academics is that individual-level output measures are observable and comparable across time and space, while similar measures are typically unavailable in other occupations. Publications are central performance metrics used to inform hiring decisions, allocate research funding, and rank scholars.

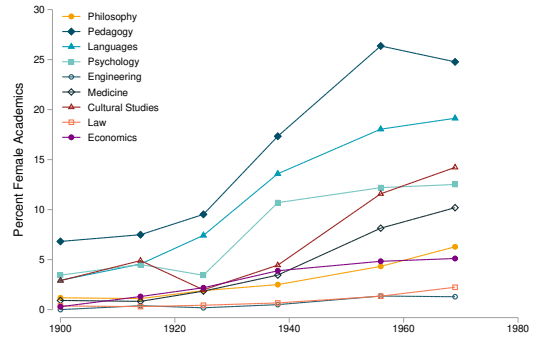
As discussed above, we do not interpret publications as direct measures of ability. Instead, they capture realized output that may reflect differences in preferences, constraints, and opportunities. Gender gaps in publications may therefore arise from multiple sources, including differential access to research resources (for example, high-quality laboratories), discrimination in peer review or in the workplace, and unequal allocation of time due to family responsibilities.

Figure 6: Percent of Female Academics by Discipline over Time

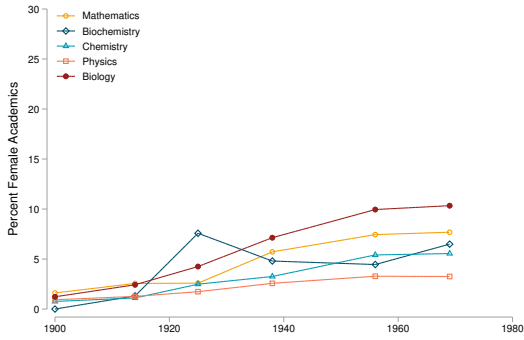
(a) *Sample 1: All universities, all disciplines, 1900-1969*



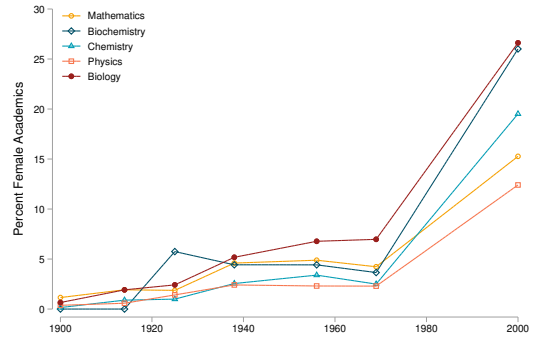
(b) *Sample 1: All universities, all disciplines, 1900-1969*



(c) *Sample 2: All universities, sciences, 1900-1969*



(d) *Sample 3: Prestigious universities, sciences, 1900-2000*



Notes: The Figure shows the percentage of female academics by discipline. Panel (a) shows female shares for all disciplines (aggregated at the level of Humanities, Social Sciences, and STEM) in all universities until 1969. Panel (b) shows female shares in nine exemplary disciplines in all universities until 1969. Appendix Figure Figure B.4 shows the remaining disciplines. Panel (c) shows female shares in the sciences (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Panel (d) shows female shares in the sciences in prestigious universities until 2000. The data were collected by the authors from *Minerva* and department websites; see Section 1 for details.

3.1 Individual-Level Publication Gaps

To estimate gender gaps in publications, we focus on academics working in the five scientific disciplines for which we have detailed publication data. For each scientist i , we observe the cohort $t(i)$ (e.g., 2000), the discipline $d(i)$ (e.g., biology), and the university $u(i)$ (e.g., Harvard). University $u(i)$ determines i 's country $c(i)$ (e.g., the United States), while i 's discipline and university ($d(i), u(i)$) determine i 's department (e.g., biology at Harvard).²³ We estimate the following regressions for Sample 2: all universities 1900-1969 and Sample 3: prestigious universities 1900-2000:

²³A small proportion of scientists have more than one affiliation in the same city and cohort, either in multiple departments of the same university or across universities. For example, the Russian-Italian chemist Maria Bakunin, who was part of a group studying the eruption of Mount Vesuvius and became the first woman to be elected to the National Academy in the physical sciences class (Ciardi and Focaccia, 2011), held appointments at the University and the Technical University of Naples. To avoid double-counting, we estimate the regressions using only one observation for each scientist and cohort. Results are very similar if we keep multiple affiliations for each scientist or if we drop scientists with multiple affiliations.

$$\begin{aligned}
\text{Pub}_{it} &= \beta_1 + \beta_2 \text{Female}_i \times 1[t(i) = 1900/14] + \beta_3 \text{Female}_i \times 1[t(i) = 1925/38] \\
&+ \beta_4 \text{Female}_i \times 1[t(i) = 1956/69] + \beta_5 \text{Female}_i \times 1[t(i) = 2000] \\
&+ \text{Experience}_{it} \beta_6 + \text{FE}(i, t) + \varepsilon_{it},
\end{aligned} \tag{1}$$

where Pub_{it} measures the number of papers that scientist i from cohort $t(i)$, discipline $d(i)$, country $c(i)$, and university $u(i)$ published in journals covered by the Web of Science. As described above, we count papers in a \pm five-year window around scientist i 's cohort $t(i)$. For example, for scientists observed in 2000, we consider papers published between 1995 and 2005. The main explanatory variables are the interactions of the female indicator Female_i with indicators for four different periods: pre-WWI (1900 and 1914 cohorts), interwar (1925 and 1938), post-WWII (1956 and 1969), and modern (2000). All regressions control for discipline-specific measures of experience, computed as the number of times a scientist is observed in the data and interacted with discipline indicators.²⁴ We estimate each regression three times, controlling for increasingly finer fixed effects:

$$\text{FE}(i, t) \equiv \begin{cases} \alpha_{t(i),d(i),c(i)} & \text{Cohort} \times \text{Discipline} \times \text{Country} & \text{or} \\ \alpha_{t(i),d(i),c(i)} + \alpha_{t(i),u(i)} & \text{Cohort} \times \text{Discipline} \times \text{Country} + \text{Cohort} \times \text{University} & \text{or} \\ \alpha_{t(i),d(i),u(i)} & \text{Cohort} \times \text{Discipline} \times \text{University} & \text{otherwise.} \end{cases} \tag{2}$$

In the baseline specification, we control for the three-way interaction $\alpha_{t(i),d(i),c(i)}$ of cohort-discipline-country fixed effects (e.g., a separate fixed effect for mathematics in the United States in 2000). These fixed effects control for differences in the number of journals (and their coverage in publication databases) across time, disciplines, and countries. The fixed effects also account for differences in publications that can be explained by women entering academia in different cohorts, disciplines, or countries. In additional specifications, we control for more stringent fixed effects, as described in (2). The most stringent set of fixed effects controls for the three-way interaction $\alpha_{t(i),d(i),u(i)}$ of cohort-discipline-university fixed effects (e.g., a separate fixed effect for biology at Harvard in 2000). To account for the potential correlation of the residual ε_{it} , we cluster the standard errors at the discipline-country level (e.g., biology in the United States).

The 1900 and 1914 cohorts of female scientists published, on average, 1.2 fewer papers than men in the full sample of all universities. The 1925 and 1938 cohorts of female scientists published 1.7 fewer papers, and the 1956 and 1969 cohorts published 2.2 fewer papers (Table 2, Sample 2, column 1, significant at the 1% level). These are substantial gaps compared to the mean number of publications, which was around 4.

²⁴We include indicators for the number of times a scientist is observed by cohort t . A scientist observed in 1956 and again in 1969 contributes two observations. For the first observation in 1956, the indicator for being observed for the first time equals 1; for the second observation in 1969, the indicator for being observed for the second time equals 1. The indicators for the third, fourth, or fifth observations are zero for this scientist. To allow for discipline-level differences in experience profiles, we interact these indicators with indicators for each discipline. Analyses restricted to each scientist's first observation, or that additionally control for years since first publication, yield similar results (Table 3). This consistency suggests that the estimated gender gaps in publications are not driven by observing women and men at different stages of their careers.

Table 2: Gender Gaps in Individual-Level Publications

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Publications			Standardized Publications		
<i>Sample 2: All Universities 1900-1969</i>						
Female (1900/14)	-1.169*** (0.435)	-0.603 (0.509)	-1.011** (0.504)	-0.212** (0.085)	-0.143 (0.103)	-0.223** (0.100)
Female (1925/38)	-1.710*** (0.418)	-1.254*** (0.335)	-1.617*** (0.411)	-0.247*** (0.022)	-0.195*** (0.031)	-0.216*** (0.040)
Female (1956/69)	-2.195*** (0.515)	-1.252*** (0.276)	-1.393*** (0.281)	-0.252*** (0.024)	-0.151*** (0.016)	-0.161*** (0.013)
Observations	68,146	68,146	68,146	68,146	68,146	68,146
R-squared	0.166	0.247	0.339	0.005	0.127	0.226
<i>Sample 3: Prestigious Universities 1900-2000</i>						
Female (1900/14)	-1.644*** (0.596)	-1.282** (0.558)	-1.721*** (0.546)	-0.277*** (0.105)	-0.245** (0.106)	-0.324*** (0.107)
Female (1925/38)	-2.643*** (0.628)	-2.079*** (0.528)	-2.458*** (0.621)	-0.348*** (0.045)	-0.267*** (0.047)	-0.294*** (0.056)
Female (1956/69)	-3.519*** (0.739)	-2.721*** (0.578)	-2.949*** (0.619)	-0.401*** (0.043)	-0.322*** (0.036)	-0.326*** (0.027)
Female (2000)	-4.294*** (0.781)	-3.853*** (0.669)	-3.567*** (0.607)	-0.217*** (0.032)	-0.201*** (0.029)	-0.187*** (0.026)
Observations	88,809	88,809	88,809	88,809	88,809	88,809
R-squared	0.245	0.256	0.279	0.018	0.057	0.112
Experience×Discipline	Yes	Yes	Yes	Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes		Yes	Yes	
Cohort×University FE		Yes			Yes	
Cohort×Discipline×University FE			Yes			Yes

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist level. Sample 2 includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Sample 3 includes scientists in prestigious universities until 2000. In columns 1-3, the dependent variable is the number of publications in a \pm five-year window around a cohort (i.e., 1995-2005 for a scientist listed in 2000). In columns 4-6, the dependent variable is publications, standardized at the cohort-discipline-country level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman, interacted with the relevant cohort(s). The regressions also control for different sets of fixed effects (see definition (2) for details) and experience-by-discipline indicators. Standard errors are clustered at the discipline-country level, with 432 clusters in Sample 2 and 184 in Sample 3. Significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Even comparing women to men within the same cohort and university (column 2), the publication gaps only shrink slightly. Note, however, that a scientist's university may be endogenous, akin to occupations in traditional Mincer regressions. Finally, in column 3, we control for cohort-discipline-university fixed effects. We thus estimate publication gaps for scientists in the same university, discipline, and cohort (e.g., Harvard biologists in 2000). Even within this restricted comparison group, we find a similar pattern of gender gaps in publications.

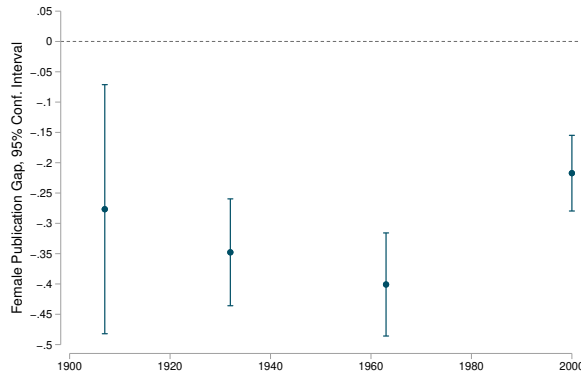
The coverage of journals in the Web of Science and the propensity to publish vary over time, across countries, and across disciplines. This affects comparisons of publication gaps because women are not equally distributed. For example, many women entered the data in later periods and worked in the United States, i.e., in periods and a country with a higher average number of publications. We therefore show alternative specifications that use standardized publications as the dependent variable. We standardize the number of publications to have a mean of 0 and a standard deviation of 1 within each country, cohort, and discipline (e.g., biology in the United States in 1969) (Table 2, Sample 2, columns

4-6).

We also show results for the sample of prestigious universities (Sample 3). This sample has two advantages. First, it enables us to extend the analysis until the year 2000. Second, in this sample, we observe nearly all universities in each of the seven cohorts. As a result, compositional changes in the sample of universities do not affect these findings.

In this sample, publication gaps are larger in absolute terms than in the unrestricted sample of universities (Table 2, Sample 3). However, because scientists at prestigious universities publish more papers on average (Table 1), the gaps are similar in percentage terms. Among prestigious universities, gender gaps in publications increased from approximately 0.28 standard deviations for the 1900 and 1914 cohorts to approximately 0.4 standard deviations for the 1956 and 1969 cohorts, before declining to 0.22 standard deviations for the 2000 cohort (see also Figure 7).

Figure 7: Gender Gaps in Publications over Time



Notes: The Figure shows gender gaps in standardized publications over time in prestigious universities (Sample 3). Gender gaps are estimated from regression (1), controlling for cohort-discipline-country fixed effects and experience-by-discipline indicators.

Robustness

The estimated gender gaps in publications are robust to alternative ways of linking papers to academics. First, results are similar if we measure publications over a \pm three-year window around i 's cohort. For example, for scientists in the 2000 cohort, we consider papers published between 1997 and 2003. As expected, the point estimates are lower because the mean number of publications is lower in a \pm three-year window (Table 3, columns 1-2). Similarly, the results are robust if we measure publications over a \pm ten-year window (Table 3, columns 3-4). Furthermore, we show that publication gaps are very similar in a sample of scientists with unique surname \times first initial \times discipline combinations in every cohort (Table 3, columns 5-6). This finding suggests that gender differences in publications are not driven by gender differences in the frequency of particular surname \times first initial pairs. Publication gaps are also similar in the sample of full professors, the academic rank that is most comparable across countries (Table 3, columns 7-8). Results are also similar in a sample that only includes each scientist once, i.e., in the cohort in which they entered the data (Table 3, columns 9-10). Finally, results are similar if we

Table 3: Individual-Level Publication Gaps (Robustness)

Dependent Variable	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)		(9)		(10)		(11)		(12)	
	±Three-Year Window				±Ten-Year Window				Unique Matches				Full Professors				First Cohort				Experience: 1st Pub.			
	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations	Publi- cations	Standard. Publi- cations
<i>Sample 2: All Universities 1900-1969</i>																								
Female (1900/14)	-0.769** (0.307)	-0.197** (0.085)	-2.622*** (0.696)	-0.254*** (0.060)	-1.163*** (0.441)	-0.205** (0.086)	-1.259*** (0.286)	-0.312*** (0.031)	-0.973** (0.459)	-0.177* (0.094)	-1.210*** (0.436)	-0.219*** (0.085)												
Female (1925/38)	-1.091*** (0.269)	-0.234*** (0.019)	-2.965*** (0.706)	-0.265*** (0.022)	-1.726*** (0.430)	-0.248*** (0.024)	-2.031*** (0.585)	-0.297*** (0.038)	-1.692*** (0.444)	-0.254*** (0.025)	-1.640*** (0.446)	-0.238*** (0.022)												
Female (1956/69)	-1.460*** (0.340)	-0.242*** (0.023)	-2.686*** (0.613)	-0.244*** (0.022)	-2.252*** (0.504)	-0.259*** (0.024)	-2.729*** (0.603)	-0.287*** (0.040)	-2.134*** (0.455)	-0.256*** (0.022)	-2.057*** (0.478)	-0.236*** (0.020)												
# Clusters (std. errors)	432	432	432	432	423	423	395	395	426	426	432	432												
Observations	68,146	68,146	68,146	68,146	60,409	60,409	35,897	35,897	50,278	50,278	68,146	68,146												
R-squared	0.156	0.005	0.170	0.005	0.166	0.008	0.200	0.017	0.169	0.010	0.244	0.064												
<i>Sample 3: Prestigious Universities 1900-2000</i>																								
Female (1900/14)	-1.108*** (0.392)	-0.259*** (0.095)	-3.420*** (1.003)	-0.314*** (0.074)	-1.628*** (0.590)	-0.271*** (0.104)	-1.541*** (0.403)	-0.350*** (0.058)	-1.492** (0.621)	-0.235** (0.115)	-2.328*** (0.887)	-0.286*** (0.106)												
Female (1925/38)	-1.699*** (0.420)	-0.329*** (0.040)	-4.617*** (1.044)	-0.380*** (0.044)	-2.650*** (0.632)	-0.347*** (0.047)	-2.917*** (0.724)	-0.404*** (0.082)	-2.538*** (0.670)	-0.350*** (0.046)	-2.641*** (0.686)	-0.337*** (0.042)												
Female (1956/69)	-2.280*** (0.500)	-0.375*** (0.043)	-4.302*** (0.875)	-0.396*** (0.043)	-3.458*** (0.744)	-0.398*** (0.046)	-4.388*** (0.953)	-0.443*** (0.055)	-3.386*** (0.723)	-0.410*** (0.053)	-2.655*** (0.665)	-0.365*** (0.037)												
Female (2000)	-2.755*** (0.501)	-0.204*** (0.032)	-7.550*** (1.367)	-0.229*** (0.034)	-4.348*** (0.777)	-0.224*** (0.031)	-0.970 (0.770)	-0.103*** (0.030)	-4.318*** (0.788)	-0.217*** (0.032)	-3.269*** (0.617)	-0.178*** (0.027)												
# Clusters (std. errors)	184	184	184	184	183	183	179	179	184	184	184	184												
Observations	88,809	88,809	88,809	88,809	78,694	78,694	37,312	37,312	75,474	75,474	88,809	88,809												
R-squared	0.241	0.017	0.255	0.018	0.239	0.022	0.309	0.038	0.243	0.021	0.406	0.100												
Experience×Discipline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes												
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes												
Years From First Publication×Discipline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes												

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist level. Sample 2 includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Sample 3 includes scientists in prestigious universities until 2000. The dependent variable is the number of publications in odd columns and standardized publications in even columns. In columns 1-2, a scientist's publications are counted in a ±three-year window around a cohort, while in columns 3-4, in a ±ten-year window around a cohort. In columns 5-12, a scientist's publications are counted in a ±five-year window. In columns 5-6, we restrict the sample to scientists whose last name - first initial - discipline - cohort combination is unique. In columns 7-8, we restrict the sample to full professors. In columns 9-10, we restrict the sample to the first cohort in which a scientist's publication is observed in the data. In columns 11-12, we control for the number of years since a scientist's first publication, interacted with the scientist's discipline. The main explanatory variable is an indicator that equals 1 if the scientist is a woman, interacted with the relevant cohort(s). The regressions control for cohort-discipline-country fixed effects (see definition (2) for details) and experience-by-discipline indicators. Standard errors are clustered at the discipline-country level. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

additionally control for each scientist’s number of years since their first publication interacted with their discipline (Table 3, columns 11-12).²⁵ These results suggest that gender gaps in publications do not stem from observing women at different career or experience stages than men. The results in Table 3 focus on the specification with cohort-discipline-country fixed effects. Adding more stringent fixed effects (as in definition (2)) does not affect the robustness of these findings.

3.2 Female Representation and the Evolution of the Gender Gap in Publications

Over the 20th century, the share of women in academia increased substantially (see part 1), potentially affecting the selection of women into academic careers. When women are severely underrepresented, entry barriers are high, and only the most exceptional women may be able to enter academia. As representation expands, these barriers may gradually decline, allowing women with a broader range of abilities to enter. At higher levels of representation, however, the increasing presence of women in academia may itself improve publishing opportunities. A larger presence of women may reduce discrimination, strengthen professional networks, and increase access to research resources. These changes can raise productivity and narrow gender gaps.

To examine whether gender gaps in publishing and hiring are systematically related, we estimate a generalized version of regression (1). This specification allows the publication gender gap to vary with the share of women in a given country-period cell:

$$\begin{aligned}
\text{Pub}_{it} = & \delta + \text{Female}_i \times 1[t(i) = 1900/38] \times \underbrace{\left[\gamma_{01} + \gamma_{11}s_{0\ell(i)}^W + \gamma_{21}(s_{0\ell(i)}^W)^2 \right]}_{g_1(s_{0\ell(i)}^W) \equiv \text{gender gap function in 1900-38}} \\
& + \text{Female}_i \times 1[t(i) = 1956/69] \times \underbrace{\left[\gamma_{02} + \gamma_{12}s_{0\ell(i)}^W + \gamma_{22}(s_{0\ell(i)}^W)^2 \right]}_{g_2(s_{0\ell(i)}^W) \equiv \text{gender gap function in 1956-69}} \\
& + \text{Female}_i \times 1[t(i) = 2000] \times \underbrace{\left[\gamma_{03} + \gamma_{13}s_{0\ell(i)}^W + \gamma_{23}(s_{0\ell(i)}^W)^2 \right]}_{g_3(s_{0\ell(i)}^W) \equiv \text{gender gap function in 2000}} \\
& + \text{Experience}_{it}\delta_{\text{exp}} + \text{FE}(i, t) + \varepsilon_{it},
\end{aligned} \tag{3}$$

where Pub_{it} measures the standardized number of papers published by scientist i in cohort $t(i)$. The variable $s_{0\ell(i)}^W$ is the share of female scientists in i ’s country-cohort pair $\ell(i)$ (e.g., the share of female scientists in the United States in 2000). We specify a period-specific gender gap as a function of the share of female scientists ($s_{0\ell(i)}^W$): $g_p(s_{0\ell(i)}^W) = \gamma_{0p} +$

²⁵For scientists who never publish in journals covered by the Web of Science, we define their first year of activity as the median year of first publication among publishing scientists in the same discipline and cohort (columns 11-12, Table 3).

$\gamma_{1p}s_{0\ell}^W + \gamma_{2p}(s_{0\ell}^W)^2$. The regression, therefore, captures how the gender gap in publications varies with the share of women in a country and period. In all specifications, we control for cohort-discipline-country (or finer) fixed effects (see definition (2) for details) and experience-by-discipline indicators.^{26,27}

We report the estimation results of regression (3) in Appendix Table C.1 and discuss them in Appendix C.2.. Figure 8 reports the estimated gender gap functions ($\hat{g}_p(s_{0\ell}^W) = \hat{\gamma}_{0p} + \hat{\gamma}_{1p}s_{0\ell}^W + \hat{\gamma}_{2p}(s_{0\ell}^W)^2$) against the share of female scientists across country-cohort pairs. Each dot in the figure represents the estimated gender gap for period p as a function of the share of female scientists in a country-cohort pair (e.g., the United States in 2000).

We uncover a U-shaped relationship between the gender gap in publications and the share of female scientists (Figure 8, panel (a)), which we refer to as the “gender U.” In country-cohort pairs with very low shares of female scientists, the estimated gender gap is small or even positive. As the share of female scientists increases, the gender gap in publications widens and becomes negative. However, once female representation increases above very low levels, these negative gaps narrow again.

Because we plot the difference in publications between men and women, changes in the gender gap can reflect changes in male publication output, female publication output, or both. To distinguish between these explanations, we plot standardized publications for men (panel (a)) and for women (panel (b)) separately against the share of female scientists across country-cohort cells (Figure 9).²⁸ Each circle is weighted by the relative size of the corresponding country within a cohort, and we superimpose a quadratic fit to summarize the relationship.

We find that the average number of publications by men varies little with the share of female scientists (Figure 9, panel (a)). In contrast, the average number of publications by women exhibits a U-shaped relationship with the share of female scientists within a country and cohort (Figure 9, panel (b)). This pattern suggests that the previously documented U-shaped gender gap in publications (Figure 8, panel (a)) is primarily driven by changes in women’s publication outcomes as female representation evolves.

Together, Figures 8 and 9 suggest the following plausible interpretation of the gender U: The most talented women entered academia first (selection effect), and a greater representation of women in academia was accompanied by greater publishing opportunities for women (empowerment effect).²⁹ In periods and countries with very low representation

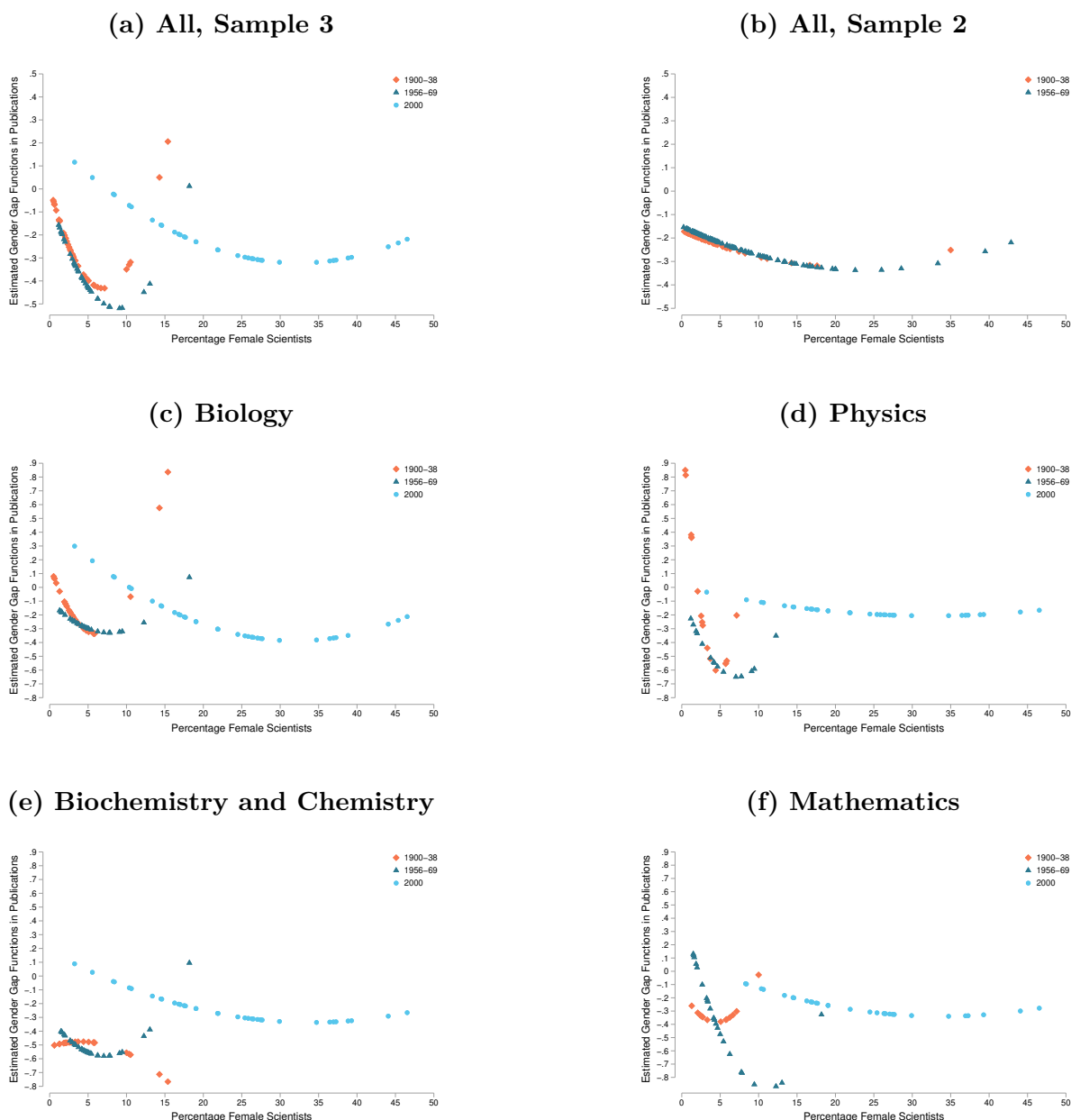
²⁶We do not separately estimate gender gap functions in publications for the 1900/1914 and 1925/1938 cohorts due to the exceedingly low numbers of female scientists in many countries in those early years. Furthermore, to reduce measurement error due to very low numbers of female scientists in some cohorts, disciplines, and countries, we compute the share of female scientists $s_{0\ell}^W$ at the level of the country-cohort pair ℓ , rather than at the finer level of the cohort-discipline-country.

²⁷We obtain qualitatively similar estimates of the gender gap function by approximating it with polynomials of degree higher than two (quadratic), such as degree three (cubic) or degree four (quartic).

²⁸To account for systematic differences in publication patterns across cohorts, disciplines, and countries, we residualize individual publication counts with respect to cohort-discipline-country fixed effects and experience-by-discipline indicators. We then compute country-cohort averages of these residualized, standardized outcomes separately by gender.

²⁹See Appendix C.2. for a discussion of this and alternative interpretations.

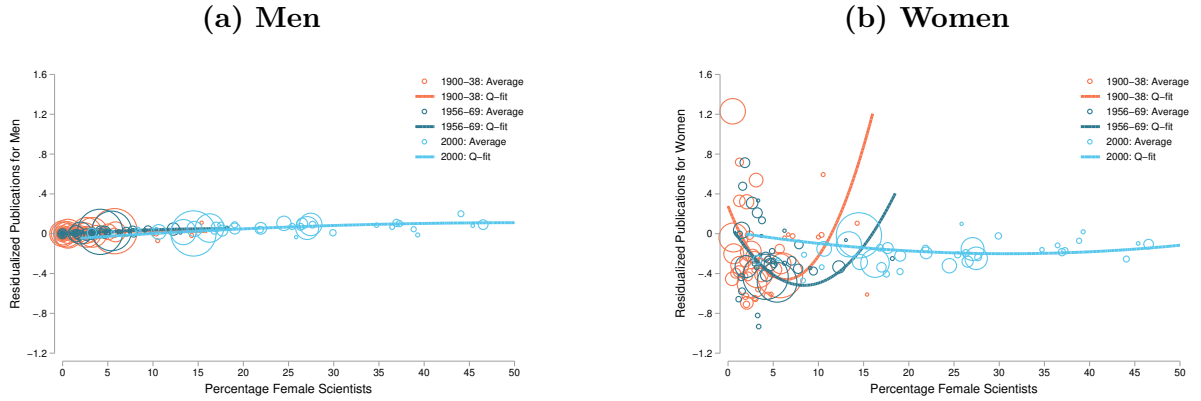
Figure 8: Gender Gaps in Publications and the Share of Women



Notes: The Figure plots the estimated gender gap in publications from regression (3) against the percentage of female scientists by country-cohort pair, in all science disciplines (panels (a)-(b)) and separately by discipline (panels (c)-(f)). Panel (a) uses the data from Sample 3, while panel (b) uses the data from Sample 2. Each dot in panels (a)-(b) represents the estimated gender gap in publications for a country-cohort pair as a function of the share of female scientists in that country-cohort pair, e.g., the United States in 2000. Each of the panels (c)-(f) reports estimated gender gap functions for a specific discipline using data from Sample 3. Each dot in panels (c)-(f) represents the gender gap in publications for a specific discipline in a country-cohort pair as a function of the share of female scientists in that country-cohort pair. All regressions control for cohort-discipline-country fixed effects and experience-by-discipline indicators.

of female scientists, selection effects are likely to dominate: only exceptionally talented women, such as Marie Curie, are able to secure positions in academia (the “Marie-Curie” periods). As representation expands and entry barriers gradually decline, women with a broader range of abilities enter academia. In this phase, the weakening of positive selection can widen gender gaps in publications if newly hired cohorts of women are, on

Figure 9: Residualized Publications by Gender and the Share of Women



Notes: The Figure plots average residualized standardized publications against the percentage of female scientists by country-cohort pair, separately for men (panel a) and for women (panel b). Both panels use data from Sample 3. We first residualize standardized publications with cohort-discipline-country fixed effects and experience-by-discipline indicators. We then average these residualized standardized publications at the country-cohort level separately for men and for women. The size of each dot in the figure corresponds to the total number of scientists in a country-cohort pair. Each panel also plots separate quadratic fits of the data for the periods 1900/38, 1956/69, and 2000.

average, less positively selected than earlier cohorts.³⁰

However, as the representation of female scientists increases, publishing opportunities may improve. A larger presence of women can facilitate the formation of successful coauthorship networks³¹ and increase women’s representation in gatekeeping roles, as more women serve as editors and referees at leading journals. This “empowerment effect” can enhance the relative output of women and contribute to a reduction in gender gaps.

In Appendix C, we show that regression equation (3) can be micro-founded with a model along the lines of Roy (1951). The model allows for (i) selection on unobservables in the hiring market, (ii) gender bias in hiring, and (iii) gender bias in the publication market. These factors contribute to the gender gap in publications through (a) indirect effects of selection and gender bias in the hiring market, and (b) direct effects of gender bias in the publication market.

Robustness

The finding of a U-shaped relationship between the gender gap in publications and the share of female scientists is robust across different samples of scientists. We show that the gender U also appears in the sample of all universities across all countries for which data are available up to 1969 (Sample 2) (Figure 8, panel (b)). This confirms that the gender U is not confined to prestigious universities but represents a broader, systematic pattern. We also confirm the gender U for most disciplines and periods (Figure 8, panels

³⁰This interpretation is consistent with Ashraf et al. (2022), who document positive selection of women into employment at a large multinational firm between 2015 and 2019, and with Mulligan and Rubinstein (2008), who find positive selection of women by ability in the broader labor market.

³¹For example, economics Nobel laureate Claudia Goldin explained: “There is one [challenge as a female academic] that wasn’t my fault, and having more women in the discipline now has allowed me to coauthor papers with the people I would like to work with, close friends. Not having female colleagues was an obstacle in terms of coauthorship [...]” See [here](#), accessed November 23, 2023.

(c)-(f)).³² This suggests that the gender U is a robust pattern that appears in most scientific disciplines and time periods.

4 Gender Gaps in Citations

The preceding analysis revealed persistent gender gaps in publications throughout the 20th century. In this section, we investigate gender gaps in scientific recognition. Specifically, we examine gender differences in citations to academic papers by estimating paper-level regressions. This approach allows us to isolate differences in recognition from the gender disparities in publications documented above.

4.1 A Novel Procedure to Predict Citations

A key challenge when estimating gender gaps in citations is that men and women may work on different topics with different citation potentials. We control for differences in citation potential using a newly developed machine-learning approach that uses paper titles to predict each paper’s expected number of citations (see Appendix D for details).³³ A similar approach can be used to study performance or pay gaps between different groups with data on, e.g., occupational task descriptions, full text of job advertisements, or performance reviews.

We first filter all non-alphanumeric characters from papers’ titles, remove common words (stopwords, e.g., “the”), and stem the words. Next, we extract all unigrams (i.e., words) and bigrams (i.e., two-word combinations) from the title of each of the N papers to obtain a paper-1,2-gram matrix \mathbf{X} with entries x_{pj} , where p denotes papers and j denotes unigrams and bigrams.³⁴ As is common in text-based machine learning, we then reweight the matrix using term-frequency inverse-document frequency (tf-idf) reweighting. This decreases the relative importance of n-grams that carry little information but appear in many papers, for example, “study” or “method.” The unigrams and bigrams then form the input for an L2-regularized regression model (ridge regression), which minimizes:

$$\min_{(\omega_j)_{j=1}^W} \left\{ \sum_{p=1}^N \left(y_p - \sum_{j=1}^W \omega_j \cdot x_{pj} \right)^2 + \lambda \sum_{j=1}^W \omega_j^2 \right\}, \quad (4)$$

where y_p is the total number of citations of paper p (standardized by country, discipline, and cohort). To reduce the importance of outliers, we winsorize citations at the 99th percentile (by discipline and cohort).³⁵ The main explanatory variables are the W indicators of the unigrams and bigrams that correspond to the respective entries of the paper-1,2-gram matrix \mathbf{X} . We additionally include indicators for the number of words in paper p ’s

³²As biochemistry is a small discipline with only a few women in the early periods, we combine chemistry and biochemistry in these regressions.

³³A pre-trained model for predicting citations is available [here](#). The model predicts the log number of citations from the titles of papers. We also provide a Python and Stata wrapper; see Schwarz (2023).

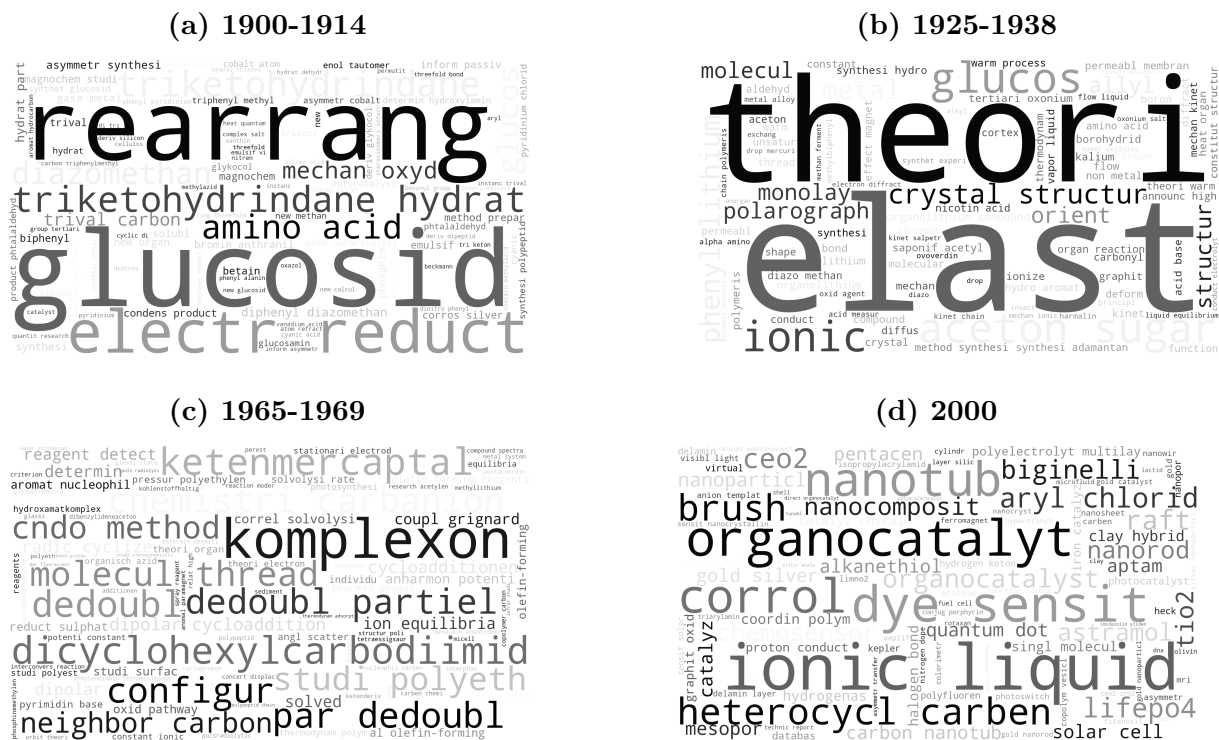
³⁴Importantly, the Web of Science translates almost all titles into English.

³⁵The results are very similar if we do not winsorize citations (Table 5).

title. To allow for differences in citation patterns over time and across disciplines, we train separate models for each of the five scientific disciplines in each of the seven cohorts. For each discipline and cohort, we choose the optimal regularization strength λ using 10-fold cross-validation. The algorithm predicts standardized citations \hat{y}_p for each paper p .

We use two approaches to estimate equation (4). For the first approach, the training sample consists of the universe of papers published by all scientists in our data. For the second approach, the training sample consists solely of papers published by men, predicting the actual citations of each paper as if it had been published by men. The model trained on all papers may provide more accurate predictions of realized citation counts. By contrast, the model trained only on male-authored papers addresses the concern that fields with substantial female contributions may be systematically undercited relative to their underlying quality, precisely because many of the papers in those areas were written by women (see Barocas and Selbst (2016) for an overview of machine-learning biases).

Figure 10: Words that Predict High Citations in Chemistry over Time



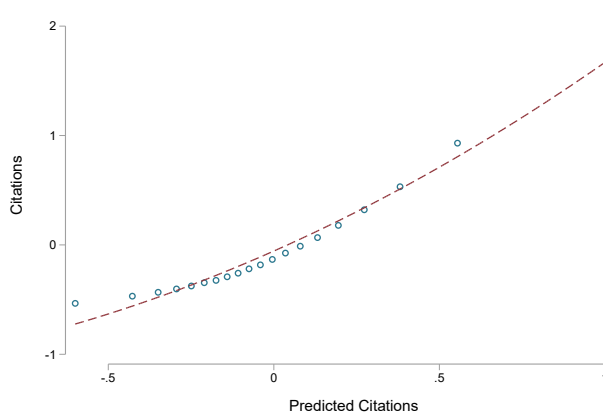
Notes: The Figure shows the unigrams and bigrams that predict the highest citations in chemistry. The n-grams are identified with an L2-regularized regression model (ridge regression) that uses unigrams and bigrams of the title as inputs, see Section 4.1 for details. A very small fraction of words in the titles are not translated into English. To improve the legibility of the word clouds, we translate them for these figures.

The model identifies intuitive relationships between words and citations. Figure 10 summarizes the unigrams and bigrams that predict high citations for chemistry and how they evolved over time.³⁶ For the 1900 and 1914 cohorts, the classifier detects n-grams related to sugar molecules (e.g., “glucos”, “glucosid”), whose chemical structures were described

³⁶To save space, we show word clouds for the model trained on papers by male scientists and combine up to two cohorts in Figure 10. However, for the regression results reported below, we implement the prediction at the cohort-discipline level.

in the 1890s by the chemistry Nobel laureate Emil Fischer. Another highly cited n-gram in this cohort is “triketohydrindane” (an alternative name for Ninhydrin), a compound discovered in 1910. For the 1925 and 1938 cohorts, the classifier identifies the words “elastic” and “thread,” which refer to the discovery of the first synthetic polymers: Nylon, developed by the DuPont chemist Wallace Carothers in 1935, and Perlon, developed by the I.G. Farben chemist Paul Schlack in 1938. For the 1956 and 1969 cohorts, the classifier detects “dicyclohexylcarbodiimid”, a widely used coupling agent for peptide synthesis based on the work of John Sheehan and George Hess in 1955. Another prominent n-gram is “CNDO method,” an abbreviation for “Complete Neglect of Differential Overlap,” one of the earliest methods in quantum chemistry developed in the 1960s by the Nobel laureate John Pople. Finally, for the 2000 cohort, the classifier detects “LiFePo4,” the chemical formula for lithium iron phosphate, a cathode material used in rechargeable batteries that was discovered in 1996. Another frequent stem in this cohort is “organocatalyt”, reflecting the growing importance of organocatalysis, a field for which Benjamin List and David MacMillan were awarded the Nobel Prize in 2021.

Figure 11: Predicted and Actual Citations



Notes: The Figure shows a binned scatter plot of the relationship between actual and predicted citations for the model trained on papers by male scientists. Actual citations are standardized at the cohort-discipline-country level. Predicted citations are estimated with an L2-regularized regression model (ridge regression) that uses unigrams and bigrams of the title as inputs, see Section 4.1 for details. The line shows a quadratic fit based on the unbinned data.

In addition to identifying intuitive relationships between words and citations, the model performs well in predicting a paper’s actual citations. Figure 11 visualizes the strong positive relationship between predicted and actual citations ($R^2 = 0.36$).³⁷ As the figure suggests, the R^2 increases further when we include a second-order polynomial of the predicted citations. In contrast, higher-order polynomials do not lead to further increases in the R^2 . We therefore control for the first- and second-degree polynomials of predicted citations, which we interact with discipline indicators in our baseline regressions.³⁸

³⁷Note that this is the within-sample R^2 . The method also performs well if we use an out-of-sample approach (see Section 4.2).

³⁸Results are very similar if we control for predicted citations either linearly or non-parametrically (Table 5).

4.2 Paper-Level Citation Gaps

We estimate citation gaps at the paper level, depending on whether papers were published by men or women. Importantly, we add our novel measures of predicted citations as regressors to control for topic-level differences in citation potential. We estimate the following paper-level regression:

$$\begin{aligned} \text{Citations}_{pt} &= \gamma_1 + \gamma_2 \text{Female}_p \times 1[t(p) = 1900/14] + \gamma_3 \text{Female}_p \times 1[t(p) = 1925/38] \\ &+ \gamma_4 \text{Female}_p \times 1[t(p) = 1956/69] + \gamma_5 \text{Female}_p \times 1[t(p) = 2000] \\ &+ \widehat{\text{Citations}}_{pt} \gamma_6 + \text{FE}(p, t) + \xi_{pt}, \end{aligned} \tag{5}$$

where p denotes a paper and $t(p)$ the cohort in which the paper was published. The dependent variable is the number of standardized citations of paper p . To reduce the importance of outliers, we winsorize citations at the 99th percentile (by discipline and cohort).³⁹ In many scientific disciplines, authors are not listed alphabetically. Instead, the first author typically contributes most of the research, while the last author is usually the senior scientist who supervises the project. We therefore define an indicator (Female_p) that equals one if either the first or the last author of paper p is female.⁴⁰ The main explanatory variables are the interactions of Female_p with indicators for four time periods: pre-WWI (1900 and 1914 cohorts), interwar (1925 and 1938), post-WWII (1956 and 1969), and modern (2000).

Importantly, we control for the first- and second-degree polynomials of predicted citations of paper p , $\widehat{\text{Citations}}_p$, interacted with discipline indicators.⁴¹ We also control for various sets of fixed effects defined at the paper level, adapting definition (2) accordingly.⁴² Note that universities, disciplines, paper titles, and, hence, predicted citations are potentially endogenous. Therefore, the estimates should be interpreted as a decomposition of citation gaps into a part that can be explained by these factors and into an unexplained part due to other biases. As before, we cluster standard errors at the discipline-country

³⁹The results are very similar if we do not winsorize citations (Table 5).

⁴⁰In mathematics, authors are typically listed alphabetically. Because many mathematics papers have only one or two authors, this definition still accurately captures authors' gender in this field. We show that results are very similar if we use alternative definitions of female-authored papers (Table 6).

⁴¹In a recent paper, Koffi (2025) propose an alternative method to assess whether papers by female authors are under-cited. This alternative approach relies on text similarity to identify papers that should have been cited. In contrast, our method controls directly for differences in the citation potential of papers and estimates overall citation gaps, rather than focusing solely on missing citations among the most similar papers. Our approach is also well-suited to settings in which the share of female authors is low, and papers by women are therefore unlikely to appear among the closest matches in similarity-based comparisons. In addition, our approach remains computationally feasible in large samples. The alternative approach by Koffi (2025) requires calculating pairwise similarities between all papers, implying that the computational burden grows quadratically with the number of papers.

⁴²Coauthors can be based in different universities and countries. We thus include separate fixed effects for any combination of cohort and university or, alternatively, cohort, discipline, and university. For example, a paper coauthored by chemists from Harvard (USA) and Cambridge (UK) has a separate fixed effect from papers coauthored by chemists only from Harvard or only from Cambridge. Accordingly, the clustering of standard errors is based on discipline and country combinations, e.g., chemistry-USA-UK in the example above.

level. Because the predicted citation control is estimated, we additionally report cluster-bootstrapped standard errors in square brackets.

Papers published by female scientists from the 1900 and 1914 cohorts received 0.12 standard deviations fewer citations than papers published by male scientists in the sample of all universities (Sample 2). Citation gaps were 0.16 standard deviations for the 1925 and 1938 cohorts and 0.13 standard deviations for the 1956 and 1969 cohorts (Table 4, Sample 2, column 1).

Table 4: Gender Gaps in Citations: Controlling for Predicted Citations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	Standardized Citations						
<i>Sample 2: All Universities, Sciences, 1900-1969</i>							
Female-First/Last Author (1900/14)	-0.118 (0.105)	-0.141* (0.084)	-0.078* (0.044)	-0.143** (0.067)	-0.062** (0.026)	-0.173*** (0.061)	-0.099*** (0.027)
		[0.137]	[0.058]	[0.118]	[0.056]	[0.116]	[0.065]
Female-First/Last Author (1925/38)	-0.155*** (0.045)	-0.149*** (0.032)	-0.106*** (0.020)	-0.121*** (0.045)	-0.106*** (0.034)	-0.103*** (0.039)	-0.092*** (0.029)
		[0.048]	[0.027]	[0.061]	[0.035]	[0.057]	[0.032]
Female-First/Last Author (1956/69)	-0.132*** (0.019)	-0.125*** (0.017)	-0.090*** (0.016)	-0.129*** (0.017)	-0.089*** (0.019)	-0.135*** (0.016)	-0.100*** (0.019)
		[0.025]	[0.02]	[0.024]	[0.023]	[0.024]	[0.024]
Observations	255,782	255,782	255,782	255,782	255,782	255,782	255,782
R^2	0.009	0.497	0.500	0.525	0.527	0.538	0.539
<i>Sample 3: Prestigious Universities, Sciences, 1900-2000</i>							
Female-First/Last Author (1900/14)	-0.172 (0.137)	-0.157 (0.117)	-0.089 (0.074)	-0.153** (0.075)	-0.063* (0.037)	-0.242*** (0.055)	-0.142*** (0.018)
		[0.122]	[0.07]	[0.095]	[0.051]	[0.084]	[0.048]
Female-First/Last Author (1925/38)	-0.109** (0.055)	-0.110*** (0.042)	-0.098*** (0.030)	-0.080 (0.055)	-0.074* (0.043)	-0.081 (0.052)	-0.073* (0.039)
		[0.056]	[0.033]	[0.066]	[0.043]	[0.062]	[0.038]
Female-First/Last Author (1956/69)	-0.164*** (0.029)	-0.140*** (0.027)	-0.113*** (0.023)	-0.132*** (0.029)	-0.102*** (0.027)	-0.142*** (0.029)	-0.111*** (0.027)
		[0.035]	[0.025]	[0.034]	[0.027]	[0.031]	[0.026]
Female-First/Last Author (2000)	-0.080*** (0.010)	-0.051*** (0.007)	-0.046*** (0.006)	-0.052*** (0.008)	-0.047*** (0.006)	-0.054*** (0.008)	-0.048*** (0.006)
		[0.008]	[0.005]	[0.009]	[0.005]	[0.009]	[0.005]
Observations	611,677	611,677	611,677	611,677	611,677	611,677	611,677
R^2	0.016	0.400	0.417	0.426	0.441	0.434	0.448
Predicted Citation Control		Yes		Yes		Yes	
Predicted Citation Control (All)			Yes		Yes		Yes
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes	Yes		
Cohort×University FE				Yes	Yes		
Cohort×Discipline×University FE						Yes	Yes

Notes: The Table shows gender gaps in citations. All regressions are estimated at the paper level. The dependent variable is the winsorized citation count, which we standardize at the cohort-discipline-country level. The main explanatory variable is an indicator that equals 1 if the paper's first or last author is a woman, interacted with the relevant cohort(s). The regressions control for different sets of fixed effects (see definition (2) for details). Additionally, the regressions in columns 2-7 control for the first- and second-degree polynomials of predicted citations. Citations are predicted based on either papers written by male scientists (columns 2, 4, and 6) or papers by all scientists (columns 3, 5, and 7). Predicted citations are based on unigrams and bigrams of papers and estimated with an L2-regularized regression model (ridge regression), see Section 4.1 for details. Standard errors are clustered at the discipline-country level, with 781 clusters in Sample 2 and 1,816 in Sample 3. We additionally report bootstrapped standard errors in square brackets. Significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Estimated citation gaps hardly change when we control for our novel measure of predicted citations (columns 2-3). This holds whether we predict citations using papers by male scientists (column 2) or by all scientists (column 3). These results indicate

that papers by female scientists were not under-cited because women worked on topics with lower citation potential, but rather because of other biases in the citation market. Citation gaps are similar if we compare papers published by scientists in the same cohort and university or even the same cohort, university, and discipline (e.g., Harvard biology in 1900) (columns 4-7).

We also find that citation gaps are comparable for papers published by female scientists from prestigious universities (Sample 3, see also Appendix Figure D.1). For the 2000 cohort, the estimated citation gaps are around 0.05 standard deviations, indicating that citation gaps still persisted at the end of the 20th century but had significantly shrunk.

Robustness

The results are similar if we use out-of-sample predictions of citations using a cross-fitting procedure (Table 5, columns 2-3). For this procedure, we divide the data into k slices. We then train the model using $k - 1$ slices. Once the model is trained, we predict citations for the left-out slice. We repeat the procedure until we obtain predictions for all k slices. The results are also robust to using alternative functional forms for the predicted citation control, i.e., controlling linearly for predicted citations (Table 5, column 4) or with 1,000 indicators for the permilles of the predicted citation distribution (column 5). The results also remain unchanged if we do not winsorize citations (column 6). Further, we show that results are robust when using citation counts, instead of standardized citations, as the dependent variable (column 7). Lastly, we show that our findings remain unchanged if we use a double machine-learning approach (Chernozhukov et al. 2018), which directly controls for the words from the titles in a version of regression (5) (column 8).⁴³

⁴³Due to the large memory requirement of the double machine-learning approach, we restrict the vocabulary to the 25,000 most frequent unigrams and bigrams (words) and do not allow the effect of words to differ by cohort and discipline.

Table 5: Citations Gaps: Controlling for Predicted Citations (Robustness)

	Baseline (1)	Out of Sample (2)	Out of Sample (3)	Linear Control (4)	Non-Parametric Control (5)	Without Winsorization (6)	Citation Count (7)	Double ML (8)
<i>Sample 2: All Universities, Sciences, 1900-1969</i>								
Female-First/Last Author (1900/14)	-0.141* (0.084) [0.137]	-0.115 (0.101) [0.135]	-0.092 (0.090) [0.121]	-0.157* (0.087) [0.144]	-0.129 (0.087) [0.143]	-0.159*** (0.043) [0.101]	-2.913** (1.419) [2.941]	-0.103* (0.055)
Female-First/Last Author (1925/38)	-0.149*** (0.032) [0.048]	-0.138*** (0.037) [0.046]	-0.136*** (0.036) [0.048]	-0.167*** (0.031) [0.048]	-0.142*** (0.031) [0.047]	-0.073*** (0.023) [0.039]	-2.973 (1.869) [2.701]	-0.123*** (0.033)
Female-First/Last Author (1956/69)	-0.125*** (0.017) [0.025]	-0.120*** (0.014) [0.02]	-0.116*** (0.013) [0.022]	-0.138*** (0.019) [0.028]	-0.122*** (0.017) [0.029]	-0.082*** (0.014) [0.021]	-7.115*** (1.045) [1.566]	-0.125*** (0.013)
Observations	255,782	255,782	255,782	255,782	255,782	255,782	255,782	255,768
R ²	0.497	0.098	0.099	0.472	0.532	0.526	0.672	
<i>Sample 3: Prestigious Universities, Sciences, 1900-2000</i>								
Female-First/Last Author (1900/14)	-0.157 (0.117) [0.122]	-0.141 (0.129) [0.139]	-0.121 (0.110) [0.145]	-0.160 (0.118) [0.122]	-0.172 (0.119) [0.123]	-0.125* (0.071) [0.09]	-4.365* (2.443) [4.375]	-0.11 (0.085)
Female-First/Last Author (1925/38)	-0.110*** (0.042) [0.056]	-0.106*** (0.041) [0.053]	-0.107** (0.041) [0.059]	-0.125*** (0.040) [0.054]	-0.117*** (0.042) [0.056]	-0.061** (0.031) [0.044]	-0.975 (2.861) [3.783]	-0.096** (0.045)
Female-First/Last Author (1956/69)	-0.140*** (0.027) [0.035]	-0.146*** (0.025) [0.032]	-0.145*** (0.024) [0.036]	-0.147*** (0.027) [0.035]	-0.136*** (0.026) [0.035]	-0.101*** (0.017) [0.027]	-8.541*** (1.775) [2.291]	-0.148*** (0.019)
Female-First/Last Author (2000)	-0.051*** (0.007) [0.008]	-0.065*** (0.008) [0.008]	-0.066*** (0.007) [0.011]	-0.056*** (0.007) [0.008]	-0.052*** (0.007) [0.008]	-0.046*** (0.007) [0.007]	-2.811*** (0.428) [0.447]	-0.067*** (0.005)
Observations	611,677	611,677	611,677	611,677	611,677	611,677	611,677	611,513
R ²	0.400	0.132	0.137	0.376	0.416	0.365	0.610	
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predicted Citations Control	Yes			Yes		Yes	Yes	
Predicted Citations Control (Out-of-Sample)		Yes						
Predicted Citations Control (All, Out-of-Sample)			Yes					
Predicted Citations (1,000 bins) FE					Yes			

Notes: The Table shows gender gaps in citations. All regressions are estimated at the paper level. The dependent variable is the winsorized citation count (except for column 6), which we standardize at the cohort-discipline-country level (except in column 7). The main explanatory variable is an indicator that equals 1 if the paper's first or last author is a woman, interacted with the relevant cohort(s). The regressions control for different sets of fixed effects (see definition (2) for details). Additionally, the regressions control for the first- and second-degree polynomials of predicted citations or for 1,000 indicators for the permilles of the predicted citation distribution interacted with discipline indicators. Predicted citations are based on unigrams and bigrams of papers and estimated with an L2-regularized regression model. Standard errors are clustered at the discipline-country level, with 781 clusters in Sample 2 and 1,816 in Sample 3. Where relevant, we additionally report bootstrapped standard errors in square brackets. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

In additional checks, we investigate alternative definitions of female-authored papers (Table 6). We show results for an indicator for any female author (column 2), female first author (column 3), and the share of women among all authors (column 4). Throughout, the estimated citation gaps remain similar. One interesting finding is that for the 2000 cohort, the estimates become insignificant for the any-female-author specification, while remaining negative for our baseline and the female-first-author specification. This suggests that the gender gap is larger for papers for which the author's gender is more salient.

Table 6: Citations Gaps: Various Definitions of Papers by Female Authors

	Baseline (1)	Female Indicator (2)	Female First-Author (3)	Female Share (4)
<i>Sample 2: All Universities, Sciences, 1900-1969</i>				
Female Paper (1900/14)	-0.141* (0.084) [0.137]	-0.144* (0.082) [0.135]	-0.133* (0.081) [0.121]	-0.135 (0.092) [0.144]
Female Paper (1925/38)	-0.149*** (0.032) [0.048]	-0.120*** (0.041) [0.046]	-0.156*** (0.042) [0.048]	-0.127*** (0.040) [0.048]
Female Paper (1956/69)	-0.125*** (0.017) [0.025]	-0.116*** (0.017) [0.02]	-0.136*** (0.023) [0.022]	-0.130*** (0.018) [0.028]
Observations	255,782	255,782	255,782	255,782
R^2	0.497	0.497	0.497	0.497
<i>Sample 3: Prestigious Universities, Sciences, 1900-2000</i>				
Female Paper (1900/14)	-0.157 (0.117) [0.122]	-0.159 (0.115) [0.139]	-0.150* (0.085) [0.145]	-0.159 (0.116) [0.122]
Female Paper (1925/38)	-0.110*** (0.042) [0.056]	-0.070 (0.054) [0.053]	-0.119** (0.053) [0.059]	-0.074 (0.056) [0.054]
Female Paper (1956/69)	-0.140*** (0.027) [0.035]	-0.139*** (0.025) [0.032]	-0.152*** (0.030) [0.036]	-0.159*** (0.030) [0.035]
Female Paper (2000)	-0.051*** (0.007) [0.008]	-0.007 (0.007) [0.008]	-0.058*** (0.010) [0.011]	-0.027*** (0.009) [0.008]
Observations	611,677	611,677	611,677	611,677
R^2	0.400	0.400	0.400	0.400
Predicted Citations Control	Yes	Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes

Notes: The Table shows gender gaps in citations. All regressions are estimated at the paper level. The dependent variable is the winsorized citation count, which we standardize at the cohort-discipline-country level. The main explanatory variables are different definitions of female-authored papers, interacted with the relevant cohort(s). In column 1, the female paper indicator equals 1 if the paper's first or last author is a woman. In column 2, the female paper indicator equals 1 if the paper has at least one female author. In column 3, the female paper indicator equals 1 if the first author of the paper is a woman. In column 4, female paper measures the share of female authors. The regressions control for different sets of fixed effects (see definition (2) for details). Additionally, the regressions control for the first- and second-degree polynomials of predicted citations. Predicted citations are based on unigrams and bigrams of papers and estimated with an L2-regularized regression model. Standard errors are clustered at the discipline-country level, with 781 clusters in Sample 2 and 1,816 in Sample 3. We additionally report bootstrap standard errors in square brackets. Significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Alternative Explanations

Apart from the possibility that women may have worked on topics with different citation potential, there are at least two additional reasons why papers by female scientists may have received fewer citations. First, women may have had fewer opportunities to write papers with coauthors (production effect). This may translate into fewer citations because coauthored papers, on average, receive more citations (e.g., Wuchty et al. 2007). Second, women may publish their papers in lower-ranked journals because of biased editors or

referees (publication effect). Such an effect has been documented for economics papers (Card et al. 2022).

Table 7: Gender Gaps in Citations: Accounting for Number of Authors and Journals

Dependent Variable	Author Nr. FE			Journal FE		
	(1)	(2)	(3)	(4)	(5)	(6)
	Standardized Citations					
<i>Sample 2: All Universities, Sciences, 1900-1969</i>						
Female-First/Last Author (1900/14)	-0.141* (0.083) [0.137]	-0.140** (0.068) [0.119]	-0.171*** (0.061) [0.116]	-0.183*** (0.069) [0.131]	-0.200*** (0.052) [0.106]	-0.241*** (0.051) [0.104]
Female-First/Last Author (1925/38)	-0.147*** (0.032) [0.048]	-0.119** (0.046) [0.062]	-0.101** (0.040) [0.057]	-0.147*** (0.034) [0.048]	-0.111** (0.046) [0.061]	-0.104** (0.042) [0.058]
Female-First/Last Author (1956/69)	-0.119*** (0.018) [0.026]	-0.122*** (0.017) [0.024]	-0.128*** (0.016) [0.024]	-0.105*** (0.019) [0.026]	-0.121*** (0.016) [0.022]	-0.126*** (0.015) [0.023]
Observations	255,782	255,782	255,782	255,782	255,782	255,782
R^2	0.498	0.526	0.539	0.520	0.546	0.558
<i>Sample 3: Prestigious Universities, Sciences, 1900-2000</i>						
Female-First/Last Author (1900/14)	-0.156 (0.112) [0.118]	-0.145* (0.075) [0.095]	-0.235*** (0.054) [0.084]	-0.237*** (0.088) [0.095]	-0.231*** (0.046) [0.069]	-0.357*** (0.034) [0.071]
Female-First/Last Author (1925/38)	-0.111** (0.043) [0.056]	-0.080 (0.057) [0.068]	-0.081 (0.054) [0.064]	-0.112*** (0.042) [0.055]	-0.074 (0.054) [0.065]	-0.091* (0.054) [0.064]
Female-First/Last Author (1956/69)	-0.134*** (0.027) [0.035]	-0.126*** (0.030) [0.034]	-0.135*** (0.029) [0.031]	-0.139*** (0.027) [0.033]	-0.143*** (0.029) [0.032]	-0.150*** (0.028) [0.03]
Female-First/Last Author (2000)	-0.037*** (0.007) [0.008]	-0.038*** (0.008) [0.009]	-0.040*** (0.008) [0.009]	-0.037*** (0.006) [0.007]	-0.039*** (0.006) [0.007]	-0.041*** (0.006) [0.007]
Observations	611,677	611,677	611,677	611,677	611,677	611,677
R^2	0.404	0.429	0.438	0.470	0.493	0.500
Predicted Citations Control	Yes	Yes	Yes	Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes		Yes	Yes	
Cohort×University FE		Yes			Yes	
Cohort×Discipline×University FE			Yes			Yes
Nr. Authors FE	Yes	Yes	Yes			
Journal FE				Yes	Yes	Yes

Notes: The Table shows gender gaps in citations. All regressions are estimated at the paper level. The dependent variable is the citation count, which we standardize at the cohort-discipline-country level. The main explanatory variable is an indicator that equals 1 if the paper's first or last author is a woman, interacted with the relevant cohort(s). The regressions control for different sets of fixed effects (see definition (2) for details). Additionally, the regressions control for the first- and second-degree polynomials of predicted citations. Predicted citations are based on unigrams and bigrams of papers and estimated with an L2-regularized regression model. In columns 1-3, we also control for fixed effects for the number of authors. In columns 4-6, we also control for journal fixed effects for the journal of paper p . Standard errors are clustered at the discipline-country level, with 781 clusters in Sample 2 and 1,816 in Sample 3. We additionally report bootstrapped standard errors in square brackets. Significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

We explore the first alternative explanation by controlling for the number of authors on each paper (i.e., a fixed effect for single-authored papers, another fixed effect for papers with two authors, and so on). Controlling for the number of authors does not affect gender gaps in citations (Table 7, columns 1-3).

Next, we examine the publication effect by including a full set of journal fixed effects (columns 4-6). The inclusion of these has little impact on the magnitude of the estimates. As with universities, disciplines, or titles, the number of coauthors and the journal are potentially endogenous. Therefore, these results should be interpreted as a decomposition of the citation gaps rather than as causal estimates.

Overall, the results indicate that gender gaps in citations do not stem from gender differences in the number of coauthors, the journals in which women publish, or, most importantly, the topics on which women work. This suggests that papers by women received fewer citations because of biases in citing behavior. Such biases could arise because women have fewer opportunities to present their work or face discrimination.

5 Gender Gaps in Promotions

In this final section, we investigate gender gaps in promotions. We focus on the sample of academics who were not already full professors when they entered the data in cohort $t - 1$.⁴⁴ We then analyze whether they are promoted to full professor by cohort t (see Appendix A.1.1. for details on the coding of promotions). Promotions to full professor are particularly important because, in all countries, full professors have unique privileges, high job security, and the highest salaries in academia. We estimate the following regression:

$$\begin{aligned} \text{Promotion Full Prof}_{it} &= \pi_1 + \pi_2 \text{Female}_i \times 1[t(i) = 1914] + \pi_3 \text{Female}_i \times 1[t(i) = 1925/38] \\ &+ \pi_4 \text{Female}_i \times 1[t(i) = 1956/69] + \pi_5 \text{Female}_i \times 1[t(i) = 2000] \\ &+ \text{FE}(i, t - 1) + v_{it}. \end{aligned} \tag{6}$$

The dependent variable $\text{Promotion Full Prof}_{it}$ is an indicator that equals 1 if academic i , who entered the data in cohort $t - 1$, was promoted to full professor by cohort t .⁴⁵ The main explanatory variables are the interactions of the indicator Female_i with indicators for the four different time periods: pre-WWI (1900 and 1914 cohorts), interwar (1925 and 1938), post-WWII (1956 and 1969), and modern (2000). The regressions include fixed effects as defined in (2), evaluated in $t - 1$. The fixed effects control, for example, for the fact that in certain time periods, disciplines, and universities, there may have been more full professor openings. Gender gaps in promotions may arise for two main reasons. First, women may have been less likely to apply for promotion, and second, their probability of promotion may have been lower, conditional on applying. We estimate the total effect. As

⁴⁴This restriction results in a smaller sample because academics who entered the data as full professors are not included in the analysis. Furthermore, all academics who entered the data in the last cohort, regardless of their rank, are also not included in the analysis.

⁴⁵We also set the indicator to 1 for academics who are listed as emeriti/emerae in t . The indicator equals 0 if the academic was not promoted to full professor by cohort t . We also set the indicator to 0 for academics who left the sample by cohort t . In unreported results, we analyze promotions in a restricted sample that conditions on observing academic i in both cohorts $t - 1$ and t . These results also indicate that women were significantly less likely to be promoted to full professor.

we do not have data on applications for full professorships, we cannot distinguish between these two possibilities.

Table 8: Gender Gaps in Promotions

	(1)	(2)	(3)	(4)
Dependent Variable:	Indicator of Promotion to Full Professor			
<i>Sample 1: All Universities, all disciplines, 1900-1969</i>				
Female (1914)	-0.127*** (0.037)	-0.128*** (0.049)	-0.088* (0.046)	
Female (1925/38)	-0.135*** (0.015)	-0.148*** (0.016)	-0.142*** (0.017)	
Female (1956/69)	-0.123*** (0.014)	-0.103*** (0.010)	-0.103*** (0.011)	
Observations	102,610	102,610	102,610	
R-squared	0.114	0.223	0.388	
<i>Sample 2: All Universities, sciences, 1900-1969</i>				
Female (1914)	-0.134 (0.129)	0.317* (0.176)	0.266 (0.169)	0.268 (0.171)
Female (1925/38)	-0.114*** (0.039)	-0.168*** (0.044)	-0.151*** (0.042)	-0.147*** (0.042)
Female (1956/69)	-0.188*** (0.015)	-0.156*** (0.020)	-0.165*** (0.020)	-0.158*** (0.020)
Std. Publications				0.048*** (0.006)
Std. Citations				0.009 (0.006)
Observations	16,649	16,649	16,649	16,649
R-squared	0.104	0.305	0.460	0.467
<i>Sample 3: Prestigious Universities, sciences, 1900-2000</i>				
Female (1914)	-0.193 (0.175)	0.810*** (0.230)	0.600*** (0.227)	0.604*** (0.229)
Female (1925/38)	-0.169*** (0.042)	-0.218*** (0.054)	-0.186*** (0.059)	-0.176*** (0.059)
Female (1956/69)	-0.222*** (0.020)	-0.238*** (0.030)	-0.240*** (0.022)	-0.230*** (0.024)
Female (2000)	-0.094*** (0.028)	-0.072*** (0.017)	-0.068*** (0.016)	-0.059*** (0.015)
Std. Publications				0.031*** (0.007)
Std. Citations				0.018*** (0.005)
Observations	12,654	12,654	12,654	12,654
R-squared	0.159	0.295	0.421	0.427
Cohort×Discipline×Country FE	Yes	Yes		
Cohort×University FE		Yes		
Cohort×Discipline×University FE			Yes	Yes

Notes: The Table shows gender gaps in the probability of promotion to full professor. Results are estimated at the academic level. Sample 1 includes academics in all disciplines and all universities until 1969. Sample 2 includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Sample 3 includes scientists in prestigious universities until 2000. The dependent variable is an indicator that equals 1 if an academic who entered the dataset in cohort $t - 1$ at a lower rank than full professor was promoted to full professor by cohort t . The main explanatory variable is an indicator that equals 1 if the academic is a woman, interacted with the relevant cohort(s). The regressions also control for different sets of fixed effects (see definition (2) for details) evaluated in $t - 1$. Standard errors are clustered at the discipline-country level, with 1,606 clusters in Sample 1, 285 in Sample 2, and 161 in Sample 3. Significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Because we observe promotion outcomes for all academics, not just those in the sciences, we can study these outcomes across the full set of disciplines. In all universities and disciplines (Sample 1), women who started their careers in the 1900 cohort were, on average, 13 percentage points less likely than men to be promoted to full professor by 1914 (Table 8, Sample 1, column 1). Because the probability of promotion to full professor was around 16% in 1914, women were about 79% less likely to be promoted. Women in the 1925 and 1938 cohorts and those in the 1956 and 1969 cohorts were around 14 and 12 percentage points (or 87% and 76%, respectively) less likely than men to be promoted to full professor by the next cohort. The large gender gap in promotions to full professor is robust to the inclusion of more stringent fixed effects. We estimate similar gender gaps in promotions if we compare women and men who started their careers in the same cohort and university or even in the same cohort, university, and discipline (Table 8, Sample 1, column 3).

In the scientist sample (mathematics, physics, chemistry, biochemistry, and biology) of all universities (Sample 2), women who started their careers in 1900 were around 13 percentage points less likely to be promoted by 1914. Women in the 1925 and 1938 cohorts and in the 1956 and 1969 cohorts were 11 and 19 percentage points less likely than men to be promoted to full professor (Table 8, Sample 2, column 1). For the cohorts after 1925, the promotion gaps are similar if we condition on more stringent fixed effects. For the 1900 cohort (i.e., those who could have been promoted by 1914), the effects turn positive if we include more stringent fixed effects. However, given the extremely low female representation in the sciences in 1900, only 16 women in all universities of the world combined could have been promoted to full professor. Comparisons of promotions within the same cohort and university (or the same cohort, university, and discipline) are thus based on only a handful of women.

The results are also similar if we estimate promotion gaps for scientists in prestigious universities (Sample 3). In this sample, we can extend the time horizon and find that promotion gaps have declined to about 6 to 9 percentage points by 2000.

In the second and third parts of the paper, we have shown that women published fewer papers and received fewer citations throughout the 20th century. To explore whether gender gaps in publications and citations affect gender gaps in promotions, we add controls for scientists' publication and citation records.⁴⁶ A one-standard-deviation better publication record increased the probability of promotion to full professor by 4.8 (Sample 2) or 3.1 (Sample 3) percentage points. A one-standard-deviation better citation record did not significantly increase the probability of promotion in Sample 2, but increased it by 1.8 percentage points in Sample 3. The small effect of citations on promotions in Sample 2 is likely driven by the fact that it has been possible to systematically measure citations only since the 1960s (Hager et al. 2024).

⁴⁶As noted above, for the first part of the 20th century, publication and citation databases do not cover the humanities and social sciences. Thus, we cannot control for publications and citations in Sample 1.

Strikingly, controlling for the publication and citation records hardly affects the estimated gender gaps in promotion to full professor.⁴⁷ The unexplained part of the promotion gap is larger than the effect of a three- to seven-standard-deviation worse publication record. This is remarkable because the true quality of women, conditional on the number of publications and citations, should be, if anything, higher in the presence of discrimination and other biases in the publication market.

Finally, we show that women were less likely to be promoted not only in the same department but also in worse or better departments (Appendix Table E.1). This suggests that women could not even achieve promotion to full professor by moving to lower-ranked departments.

6 Conclusion

This paper provides the first global account of gender gaps in academia over the 20th century. We assemble a new database covering more than half a million academics from over 7,000 universities in more than 150 countries, harmonize academic ranks and disciplines, link academics across cohorts, and connect these records to publication and citation data. These data allow us to follow women’s entry into academia, their research output, the recognition of their work, and their progression to senior positions across countries, disciplines, and time.

Four main findings emerge. First, women were sharply underrepresented in academic positions throughout the 20th century. Although female representation increased over time, women remained a small minority in most countries and disciplines, especially in the sciences and among full professors. Second, women published fewer papers than men. The evolution of this gap followed a pronounced U-shape: it was relatively small when very few women entered academia, widened as access expanded, and narrowed again as female representation increased further. This pattern suggests that barriers to entry and gaps in subsequent research output were closely connected. Third, papers by women received fewer citations. Using a new measure of predicted citations based on paper titles, we show that this citation gap cannot be explained by women working on topics with lower citation potential. It is also not explained by differences in the number of coauthors or by the journals in which women published. Fourth, women were substantially less likely to be promoted to full professor. This promotion gap remains large even after controlling for publication and citation records, suggesting that measured academic output does not account for women’s slower career progression.

These findings indicate that gender inequality in academia was not confined to a single stage of the career. Women faced barriers to entry, produced fewer publications

⁴⁷In unreported results, we control more flexibly for publications and citations by including indicators for various percentiles of the distributions of publications and citations or by including polynomials of the standardized numbers of publications and citations. The results are similar to those reported in Table 8.

once inside academia, received less recognition for their work, and advanced more slowly to senior positions.

The paper also speaks to a broader question in economics: how much talent societies leave unused when access to high-skilled professions is restricted. Academia is a central institution for the production of ideas, which are a key input into scientific and economic progress. The exclusion and slower advancement of women, therefore, had costs that went beyond the careers of individual scholars. It likely changed which questions were asked, which discoveries were made, and how quickly knowledge advanced.

Our evidence also has implications for contemporary debates. The sharp narrowing of some gaps over time demonstrates that gender inequality in academia is not immutable. Yet the persistence of gaps in hiring, citations, and promotions, even after accounting for observable characteristics, shows that formal access alone is insufficient to ensure equal opportunity. Understanding where barriers arise, how they interact across career stages, and which institutions reduce them remains essential for designing policies that allow the best ideas to emerge, regardless of who produces them.

References

- Abramitzky, R., L. Greska, S. Pérez, J. Price, C. Schwarz, and F. Waldinger (2024). Climbing the ivory tower: How socio-economic background shapes academia. Technical report, National Bureau of Economic Research.
- Albanesi, S., C. Olivetti, and M. J. Prados (2015). Gender and Dynamic Agency: Theory and Evidence on the Compensation of Top Executives. In *Gender in the Labor Market*, pp. 1–59. Emerald Group Publishing Limited.
- Altonji, J. G. and R. M. Blank (1999). Race and Gender in the Labor Market. *Handbook of Labor Economics* 3, 3143–3259.
- Ash, E. and S. Hansen (2023). Text algorithms in economics. *Annual Review of Economics* 15, 659–688.
- Ashraf, N., O. Bandiera, V. Minni, and V. Quintas-Martínez (2022). Gender Roles and the Misallocation of Labour Across Countries. *mimeo LSE*.
- Azmat, G. and R. Ferrer (2017). Gender Gaps in Performance: Evidence from Young Lawyers. *Journal of Political Economy* 125(5), 1306–1355.
- Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017). Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. *American Economic Review* 107(3), 714–747.
- Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017). Does the Gender Composition of Scientific Committees Matter? *American Economic Review* 107(4), 1207–1238.
- Barocas, S. and A. D. Selbst (2016). Big Data’s Disparate Impact. *Calif. L. Rev.* 104, 671.
- Becker, S. O., V. Lindenthal, S. Mukand, and F. Waldinger (2024). Scholars at Risk: Professional Networks and High-Skilled Emigration from Nazi Germany. *American Economic Journal: Applied Economics* 16(3), 1–45.
- Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *The Quarterly Journal of Economics* 134(2), 647–713.
- Bertrand, M. (2011). New Perspectives on Gender. In *Handbook of Labor Economics*, Volume 4, pp. 1543–1590. Elsevier.

- Bertrand, M. and E. Duflo (2017). Field Experiments on Discrimination. *Handbook of Economic Field Experiments 1*, 309–393.
- Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics 2*(3), 228–255.
- Bertrand, M. and K. F. Hallock (2001). The Gender Gap in Top Corporate Jobs. *ILR Review 55*(1), 3–21.
- Blau, F. D. and L. M. Kahn (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature 55*(3), 789–865.
- Card, D., S. DellaVigna, P. Funk, and N. Iriberry (2022). Gender Differences in Peer Recognition by Economists. *Econometrica 90*(5), 1937–1971.
- Card, D., S. DellaVigna, P. Funk, and N. Iriberry (2023). Gender Gaps at the Academies. *Proceedings of the National Academy of Sciences 120*(4), e2212421120.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal 21*(1), C1–C68.
- Ciardi, M. and M. Focaccia (2011). Maria Bakunin (1973-1969). In *European Women in Chemistry*. Wiley Online Library.
- Curie, E. (1938). *Madame Curie*. Paris, France: Gallimard.
- De la Croix, D., F. Docquier, A. Fabre, and R. Stelter (2023). The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000–1800). *Journal of the European Economic Association*, jvad061.
- Dossi, G. (2024). Race and science. *mimeo EIEF*.
- Dupas, P., A. Handlan, A. S. Modestino, M. Niederle, M. Seré, H. Sheng, J. Wolfers, and S. D. Collective (2026). Gender differences in economics seminars. *American Economic Review 116*(2), 749–789.
- Einiö, E., J. Feng, and X. Jaravel (2019). Social push and the direction of innovation. *Available at SSRN 3383703*.
- Gayle, G.-L., L. Golan, and R. A. Miller (2012). Gender Differences in Executive Compensation and Job Mobility. *Journal of Labor Economics 30*(4), 829–872.
- Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica 87*(4), 1307–1340.
- Goldin, C. (1989). Life-Cycle Labor-Force Participation of Married Women: Historical Evidence and Implications. *Journal of Labor Economics 7*(1), 20–47.
- Goldin, C. (1990). *Understanding the Gender Gap: An Economic History of American Women*. Oxford, GB: Oxford University Press.
- Goldin, C. and L. F. Katz (2016). A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation. *Journal of Labor Economics 34*(3), 705–746.
- Hager, S., C. Schwarz, and F. Waldinger (2024). Measuring Science: Performance Metrics and the Allocation of Talent. *American Economic Review 114*(12), 4052–4090.
- Hansen, S., M. McMahon, and A. Prat (2018). Transparency and deliberation within the fomc: A computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.
- Hengel, E. (2022). Publishing while female: Are women held to higher standards? evidence from peer review. *The Economic Journal 132*(648), 2951–2991.
- Hill, R. and C. Stein (2021). *Race to the Bottom: Competition and Quality in Science*. Northwestern University and UC Berkeley.

- Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of political economy* 124(5), 1423–1465.
- Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The Allocation of Talent and US Economic Growth. *Econometrica* 87(5), 1439–1474.
- Huang, J., A. J. Gates, R. Sinatra, and A.-L. Barabási (2020). Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines. *Proceedings of the National Academy of Sciences* 117(9), 4609–4616.
- Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science. *The Quarterly Journal of Economics* 133(2), 927–991.
- Jones, C. I. (1995). R&D-Based Models of Economic Growth. *Journal of Political Economy* 103(4), 759–784.
- Killingsworth, M. R. and J. J. Heckman (1986). Female Labor Supply: A Survey. *Handbook of Labor Economics* 1, 103–204.
- Koffi, M. (2025). Innovative ideas and gender (in) equality. Technical Report 7.
- Kricheldorf, H. R. (2014). *Erkenntnisse und Irrtümer in Medizin und Naturwissenschaften*. Berlin, Germany: Springer-Verlag.
- McGrayne, S. B. (1998). *Nobel Prize Women in Science: their Lives, Struggles, and Momentous Discoveries*. New York, N.Y.: Birch Lane Press, Carol Publishing Company.
- Mitchell, B. R. (1982). *International Historical Statistics: Africa and Asia*. London, GB: The Macmillan Press LTD.
- Mitchell, B. R. (1983). *International Historical Statistics: The Americas and Australasia*. London, GB: The Macmillan Press LTD.
- Mitchell, B. R. (1993). *International Historical Statistics: Europe 1750-1988*. New York, N.Y.: Stockton Press.
- Mokyr, J. (1992). *The Lever of Riches: Technological Creativity and Economic Progress*. New York: Oxford University Press.
- Moser, P. and S. Kim (2021). Women in Science. Lessons from the Baby Boom. Technical report, National Bureau of Economic Research.
- Mulligan, C. B. and Y. Rubinstein (2008). Selection, Investment, and Women’s Relative Wages over Time. *The Quarterly Journal of Economics* 123(3), 1061–1110.
- Nature (1930). Index Generalis: Annuaire Général des Universités, Grandes Écoles, Académies, Archives, Bibliothèques, Institute Scientifiques, Jardins Botaniques et Zoologiques, Musées, Observatoires, Stés Savantes. *Nature* 126, 126–127.
- Olivetti, C. and B. Petrongolo (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics* 8, 405–434.
- Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy* 94(5), 1002–1037.
- Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy* 98(5), S71–S102.
- Ross, M., B. Glennon, R. Murciano-Goroff, E. Berkes, B. Weinberg, and J. Lane (2022). Women are Credited Less in Science than are Men. *Nature* 608(7921), 135–145.
- Rossiter, M. W. (1982). *Women Scientists in America: Struggles and Strategies to 1940*, Volume 1. Baltimore, MD: JHU Press.
- Roussille, N. (2021). The Central Role of the Ask Gap in Gender Pay Inequality. *University of California, Berkeley*.
- Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers* 3(2), 135–146.
- Santamaría, L. and H. Mihaljević (2018). Comparison and Benchmark of Name-to-Gender Inference Services. *PeerJ Computer Science* 4, e156.

- Sarsons, H. (2017a). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. *mimeo UBC*, 141–145.
- Sarsons, H. (2017b). Recognition for Group Work: Gender Differences in Academia. *American Economic Review* 107(5), 141–145.
- Sarsons, H., K. Gërkhani, E. Reuben, and A. Schram (2021). Gender Differences in Recognition for Group Work. *Journal of Political Economy* 129(1), 101–147.
- Schwarz, C. (2023). Estimating Text Regressions Using `txtreg_train`. *The Stata Journal* 23(3), 799–812.
- Shanghai Ranking (2020). Academic Ranking of World Universities 2020. <https://www.shanghairanking.com/rankings/arwu/2020>.
- Shepley, C. F. (2008). *Movers and Shakers, Scalawags and Suffragettes: Tales from Bellefontaine Cemetery*. St.Louis, MO: Missouri History Museum.
- Sinha, A., Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. Hsu, and K. Wang (2015). An Overview of Microsoft Academic Service (MAS) and Applications. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 243–246.
- Truffa, F. and A. Wong (2022). Undergraduate Gender Diversity and Direction of Scientific Research. *Work. Pap., Stanford Univ., Stanford, CA*.
- Vaughn, G. F. (2004). Katharine Coman: America’s First Woman Institutional Economist and a Champion of Education for Citizenship. *Journal of Economic Issues* 38(4), 989–1002.
- Waldinger, F. (2012). Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *Review of Economic Studies* 79(2), 838–861.
- WHED (2024). World higher education database (whed) portal. Accessed on 2024-06-17.
- Wigner, E. P. (1972). Obituaries - Maria Goeppert Mayer. *Physics Today* 25(5), 77.
- Wuchty, S., B. F. Jones, and B. Uzzi (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science* 316(5827), 1036–1039.

Online Appendix

The Appendix presents further details on the data collection, additional results, and robustness checks:

- Appendix A provides details on the data collection.
- Appendix B shows additional results on hiring gaps.
- Appendix C presents additional details on the Roy model.
- Appendix D provides additional details and results on citation gaps.
- Appendix E shows additional results on promotion gaps.

A Further Details on Data

A.1. Enhancements of Faculty Roster Data

A.1.1. Additional Information on the Coding of Academic Ranks

Minerva and the university websites report academic ranks for most academics. The ranks are reported either in the original language (e.g., maître de conférence) or are translated into English or German. Overall, the sources report almost 4,000 different combinations of countries and ranks. We recode them at the country level, because certain labels of ranks do not necessarily describe the same academic rank across countries. For example, a lecturer in the British system has a higher academic rank than a lecturer in the U.S. system. We classify all positions into the following categories: professorial admin position (e.g., dean or head of department), full professor, associate professor, assistant professor, honorary professor, clinical faculty, visiting professor, teaching position, Emerita/us, Emerita/us associate professor, Emerita/us assistant professor. In a few cases, the sources list academics who hold different academic ranks under only one heading (e.g., a joint heading of “associate and assistant professors” instead of listing “associate professors” and “assistant professors” under separate headings). In these cases, we assign the highest listed rank to each academic.

In many academic systems (e.g., in Germany and Switzerland), young researchers climb the academic ladder by substituting (filling in) for full professors for several years, before subsequently being promoted to a full professorship. For these countries, we code substitute professors as assistant professors.

For the analysis of promotions, we recode different positions into four ranks:

1. professors (comprising the categories: professorial admin position, full professor, and Emerita/us)

2. associate professors (comprising the categories: associate professor and Emerita/us associate professor)
3. assistant professors (comprising the categories: assistant professor, Emerita/us assistant professor, and clinical faculty)
4. lower-ranked positions (comprising the categories: teaching position and research position).

Promotion to Full Professor

We classify as promoted to full professor any academic who enters the data in cohort $t - 1$ at ranks 2, 3, or 4 and is promoted to rank 1 by cohort t . As the exact rank of honorary and visiting professors is not clearly defined, and their number is very low, we do not consider them for the results on gender gaps in promotions (Section 5).

A.1.2. Additional Information on the Coding of Disciplines

As described in the main text, we manually recode over 100,000 different specializations (e.g., “Advanced Reactor Theory and Quantum Theory” or “Physique des particules élémentaires”) into 36 disciplines (e.g., physics, economics, law, theology, or history). The definition of disciplines follows the classification of academic disciplines according to the German Statistical Agency (see [here](#) for details).

Some academics report multiple disciplines. When we match these academics to publications, we use the discipline they report first. For academics observed across multiple cohorts who report different disciplines across cohorts, we assign the discipline that is most frequently reported.

A few academics are reported without specializations, but some of them are reported as members of certain departments: e.g., “department of architecture” or “medical school.” If the department coincides exactly with one of the disciplines (e.g., architecture or medicine), we assign the discipline on the basis of the department.

A.1.3. Identifying Academics with Multiple Appointments within a City

We identify academics with multiple appointments within a city by hand-checking all academics with duplicate surnames. We then determine whether two entries refer to the same academic based on the first name, specialization, academic rank, and title. If an academic holds two appointments, we harmonize the first name and collapse the two entries into a single observation. The resulting observation then contains the information on all appointments and specializations of an academic within a city.⁴⁸

⁴⁸In very rare cases, academics with the same surname, first name, and discipline are observed in the same cohort but in different cities. It is often impossible to determine whether this is the same academic holding multiple appointments. We, therefore, treat such observations as two separate observations. We show that all results are very similar in a sample of academics with unique combinations of surname, first initial, and discipline.

A.1.4. Linking Academics Across Cohorts

As described in the main text, we link academics across cohorts. This linking over time is crucial to analyzing promotions. Our linking procedure allows for the possibility that academics report slightly different first names in two adjacent cohorts. Such variation in first names occurs because of five main reasons:

1. Universities sometimes report first names with slight variations across cohorts. For example, the University of Leipzig reported the geographer Joseph Partsch as *Joseph* Partsch in the 1914 cohort but as *Josef* Partsch in the 1925 cohort.
2. In certain cohorts, some universities only report their professors using an abbreviated first name plus the surname. In other cohorts, they report professors with their full first name. For example, the University of Berlin theologian Johannes Witte was reported as *Johs. Witte* in the 1925 cohort but as *Johannes Witte* in the 1938 cohort.
3. In certain cohorts, some universities only report their professors using initials plus the surname. In other cohorts, they report professors with their full first name. For example, the University of Chicago botanist Henry Chandler Cowles was reported as *Henry C. Cowles* in 1914 but as *H. C. Cowles* in 1925.
4. Original names are Germanized or Anglicized for some individuals in certain cohorts. For example, the Hungarian mathematician Gusztáv Rados was listed as *Gusztáv Rados* in 1925 but as *Gustav Rados* in 1938.
5. Variations in first names may, in rare cases, also occur because of typos either introduced by the publishers of *Minerva*, by typing mistakes of the research assistants, or by OCR errors that were not spotted by the research assistants.

Linking Academics within Departments

We first link academics who remain in the same department between cohorts t and $t + 1$. In a first step, we obtain potential links by merging academics from discipline d , country c , and university u , in cohort t , to academics from the same discipline d , same country c , and same university u in cohort $t + 1$ based on the academic's surname and first initial. In a second step, we process these potential links as follows (Note: all potential links have identical surnames, initials, disciplines, and universities and, hence, cities and countries):

1. If the recorded first name is identical in both cohorts, we classify these academics as linked (in some cases, the recorded first name may consist of one or more initials only in both cohorts).
2. If the information on the first name differs across the two cohorts, research assistants examine each potential link and decide whether the academics are the same. For example, the data contain the following academics in 1925 and 1938:

Table A.1: Examples within Department Merge

	Cohort	Surname	First Name	University	Country	Field
1	1925	Randall	Harrison Mc Allister	University of Michigan	USA	Physics
2	1938	Randall	Harrison McAllister	University of Michigan	USA	Physics
3	1925	Cerban	Albert	University of Bukarest	Romania	Law
4	1938	Cerban	Alexandru	University of Bukarest	Romania	Law

The research assistants would classify lines 1 and 2 as linked (note the small difference in the first name, otherwise this academic would already be linked in step 1). In contrast, the research assistants would not classify lines 3 and 4 as linked (even though they have the same first initial). To decide whether two lines are linked, the research assistants only allow for minor differences in the spelling of the first name, such as Harrison Mc Allister and Harrison McAllister (lines 1 and 2).

Linking Academics across Departments in the same Country

Second, we link academics who remain in the same country but change departments between two cohorts. In the first step, we obtain potential links by merging academics from discipline d , country c , cohort t to academics from the same discipline d , same country c , but cohort $t + 1$ based on the academic’s surname, the first initial, their discipline, and country. Hence, all potential links that we consider have identical surnames, initials, disciplines, and countries, but they are listed in different universities (in cohort t and cohort $t + 1$) in the same country, and the first name is not necessarily identical.⁴⁹ We then process the potential links as follows:

1. If the recorded first name is identical in both cohorts, we classify these academics as linked (in some cases, the recorded first name may consist of one or more initials only in both cohorts).
2. If the information on the first name differs across the two cohorts, research assistants examine each potential link and decide whether the academics are the same. To decide whether a potential link is correct, the research assistants use the following rules:
 - (a) If there are only minor spelling differences in the first name, the research assistant classifies the potential link as correct (see lines 1 and 2 in Appendix Table A.2)

⁴⁹A small number of universities change country over the time period we consider in our analysis. For example, the University of Strasbourg is listed as a German university in 1900 and 1914, but as a French university from 1925 onward. Hence, the within-country link for the University of Strasbourg considers academics who moved from or to other German universities between 1900 and 1914, as well as between 1914 and 1925. It also considers academics who moved from or to other French universities between 1914 and 1925 and all following cohorts. The moves from Germany to Strasbourg or from Strasbourg to France between 1914 and 1925 (when the university changed country) are considered in the cross-country link that we describe below.

- (b) If all initials of the first name are identical and if the first name contains more than one initial (even if the first name differs e.g., because the academic is listed with the full first name in one cohort and with initials in the other cohort) the potential link is classified as correct (see lines 3 and 4 in Appendix Table A.2)
- (c) If only one initial is reported for one cohort, but a full first name in the other cohort, the research assistants Google the relevant academic. If the research assistants find online biographical information that confirms that the academic was indeed employed at university u in the year corresponding to cohort t and then moved to university u' before the year corresponding to cohort $t + 1$, the potential link is classified as correct. For example, K(arl) Röder (see lines 5 and 6 in Appendix Table A.2) could be found online and his [Wikipedia entry](#) states that:

“In 1924, Röder went to the Technical University of Stuttgart as a full professor of machine parts, gear mechanics and machine science. In 1926 he moved to the TH Hanover on the chair of steam engines...”
(translated with Google Translate).

In contrast, if the research assistants cannot find enough biographical information, such as for T(ito) Tosi (lines 7 and 8 in Appendix Table A.2), they classify the potential link as incorrect.

Table A.2: Within Country Merge

	Cohort	Surname	First Name	University	Country	Field
1	1925	vilinskij	sergej g.	Masarykova Universita	Czechoslovakia	Languages
2	1938	vilinskij	sergij g.	Masarykova Universita	Czechoslovakia	Languages
3	1925	jones	o. t.	University of Manchester	UK	Geology
4	1938	jones	owen thomas	University of Cambridge	UK	Geology
5	1925	roder	k.	Technische Hochschule Stuttgart	Germany	Engineering
6	1938	roder	karl	Technische Hochschule Hannover	Germany	Engineering
7	1925	tosi	t.	Universita degli Studi Messina	Italy	Languages
8	1938	tosi	tito	Universita degli Studi di Firenze	Italy	Languages

Linking Academics across Countries

Third, we link academics who move across countries. In the first step, we obtain potential links by merging academics with the same surname, first initial, and discipline d in cohort t to academics with the same surname, first initial, and discipline d in cohort $t + 1$ who are listed in two different countries. Thus, all potential links that we consider have identical surnames and disciplines, but are listed in different countries, and the first name is not necessarily identical.

To rule out false positives, all potential links are confirmed by extensive manual online searches. If the research assistants find online biographical information that confirms that the academic was employed by university u in country c in the year corresponding to cohort t and then moved to university u' in country c' before the year corresponding to cohort $t + 1$, the potential link is classified as correct.

A.1.5. Increasing the Share of Academics with Full First Names

For most academics, we infer their gender on the basis of their first name and their country.⁵⁰ The raw data report full first names for about 77% of academics. We increase the share of academics with full first names in two ways. First, we use information on the same academic from a different cohort (see Appendix A.1.4. for details on how we link academics across cohorts). For example, the University of Chicago botanist Henry Chandler Cowles was reported as Henry C. Cowles in 1914 but as H. C. Cowles in 1925. After linking the observations, we adjust the first name in 1925 to Henry C and can therefore code the gender of Cowles in 1925.

Second, we hand-check around 60,000 academics who are reported with initials in all cohorts. For this step, research assistants Google the initial(s), surname, discipline, and university to find online records for the respective academics. If they find a record, they adjust the first name to include as much information as possible.

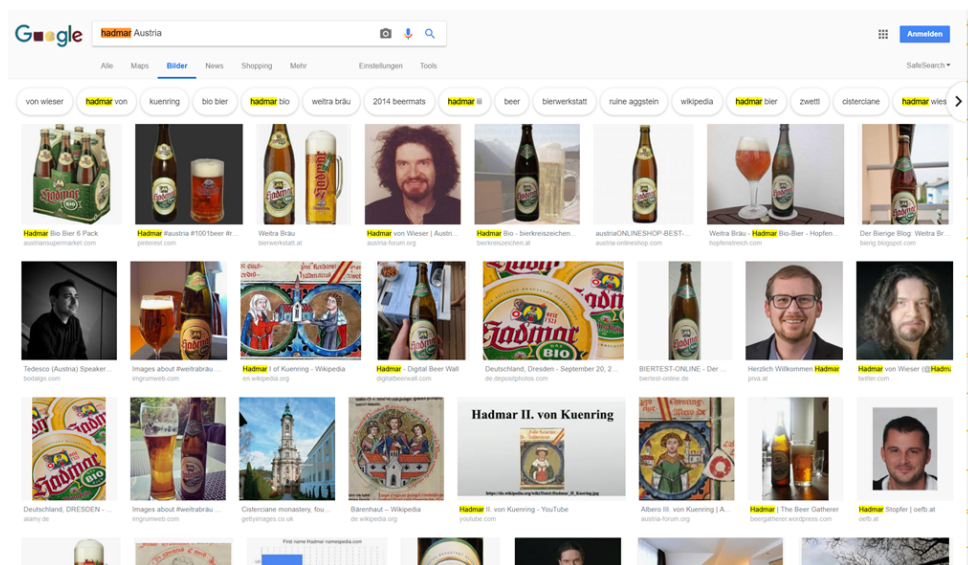
These enhancements increase the share of academics with full first names from around 77% to around 81%. Note, however, that none of the results in this paper depend on these enhancements.

A.2. Additional Information on Coding Gender

A.2.1. Example Google Picture Search

As described in the main text, one of the steps to identify the gender of academics relies on a Google image search for some first-name-country combinations. Appendix Figure A.1 shows an example of the output of the Google image search for “Hadmar Austria.”

Figure A.1: Example Google Picture Search for Assignment of Gender



Notes: The Figure shows an example of the Google image search. We apply this search to increase the share of first names by country combinations that can be assigned as male or female. The Google image search is used if gender-api.com and the hand-coding of research assistants cannot assign gender to a first name by country.

⁵⁰For some academics, we can use the information on gender from the way that academics are listed in *Minerva* (e.g., as Miss or Mlle.) or from their website (e.g., from pictures or personal pronouns).

A.2.2. Hand-Checking Gender Coding

As described in the main text, in the last step of assigning gender, we hand-check individual academics who appear misclassified. Such misclassifications occur mostly because the predominant gender of a first name - country combination changes over time.⁵¹ For example, French academics with the first name Camille were predominantly male in the early part of the 20th century. In contrast, during the latter half of the century, many French academics with the first name Camille were female. We hand-check such cases as follows: first, we identify first name - country combinations with the potential of misclassification (e.g., Camille in France); Second, research assistants Google the actual academic and try to establish their gender. For example, for the French biologist Camille Sauvageau, they find an entry in the Proceedings of the Linnean Society of London (from 1937) which says: “Camille Sauvageau (1861-1936), Foreign Member of the Society, was born in Angers on 12 May 1861. *He* studied at Montpellier...” (see [here](#) for details). This allows us to identify him as male.

A.3. Details on Merging Web of Science with Faculty Rosters

We improve the match between the faculty rosters and the Web of Science by homogenizing author names and by enriching and homogenizing the address data in the Web of Science using the multi-step procedures described below.

A.3.1. Homogenizing Author Names

Because the link between the faculty rosters and the Web of Science relies on name information, we first homogenize author names across the two sources.

The Web of Science lists a string variable corresponding to the name of each author of the paper. For simplicity, we refer to this variable as “full scientist name.”⁵² For papers published during and after the 1970s, the full scientist name reports the scientist’s name as printed on the original article, e.g., “Whish, William J. D.” For papers published before the 1970s, however, the full scientist name abbreviates the first name(s) of a scientist to the corresponding initial(s), e.g., “Whish, W. J. D.” To improve the quality of the merge between the Web of Science and the faculty rosters, we homogenize names by processing them as follows:

1. We remove titles such as “Jr.”, “Dr.”, or “Prof.” from the name.
2. We separate the full scientist name into two variables, the scientist’s surname and the scientist’s first name(s) or initials. The standard format of the full scientist name is “surname, first name(s)” and we rely on the position of the comma “,” to separate the surname and the first name(s).

⁵¹Gender-api.com (or any other professional solution that identifies the gender of first names by country) does not have enough underlying data to allow the gender prediction to differ by time periods.

⁵²In very rare cases, the Web of Science lists multiple coauthors with identical surnames and initials. Manual checks confirm that most of these are mistakes that occurred in the data entry by the Web of Science. We, therefore, keep only one of such observationally equivalent coauthors.

3. We remove noble titles, e.g., “Della”, “Op Den”, “Von Der”, or “Viscount.”
4. We extract initial(s) from the scientist’s first name(s).
5. We further extract the first of the initials from the list of initials obtained in the previous step.

A.3.2. Preparing Addresses in the Web of Science and the Faculty Rosters

The link between the faculty rosters and the Web of Science also relies on information about an academic’s country and city. We improve the quality of this link by following the steps below.

Enriching the Address Data from Web of Science with Address Data from Microsoft Academic Graph

Sometimes, the Web of Science does not report scientists’ addresses, even though the original paper actually lists an address. In some of these instances, an alternative database, Microsoft Academic Graph (MAG), contains the relevant address information. We therefore enrich the affiliations as reported by the Web of Science with information from MAG.⁵³ We match the information from MAG to the Web of Science as follows:

1. We match the scientist-paper observations that are unique in i) the journal name, ii) the year of publication, iii) the last word of the scientist’s surname, and iv) the first page of the paper.
2. We then match the scientist-paper observations that are unique in i) the journal name, ii) the year of publication, iii) the initial of the scientist’s surname, and iv) the first page of the paper.
3. We finally match the remaining scientist-paper observations that are unique in i) the journal name, ii) the year of publication, iii) the last word of the scientist’s surname, and iv) the first and last words of the paper title.

Expanding Addresses within Journals and Years

We also increase the share of papers with addresses by using information from papers published by the same author in the same year and journal. For example, Ball JM published a paper in 1900, vol. 34, January-June issue of the *Journal of the American Medical Association* for which we observe the city “St. Louis” and the country “USA.” Ball JM then published another paper in 1900, vol. 35, July-December issue of the *Journal of the American Medical Association*, for which we do not observe an affiliation. We then assign the city and country information, “St. Louis, USA,” from the first paper to the second paper.

⁵³MAG is a publicly available database of academics, their papers, and citations (see Sinha et al. 2015 for details). While MAG is freely available, the coverage until 1950 is much less comprehensive than the Web of Science. We, therefore, use the Web of Science as the main source for publications and citations.

Processing Addresses with Google Maps

Because the link between the faculty rosters and the publication and citation data relies on an academic’s country and city, we harmonize country and city information across the faculty rosters and the Web of Science.

This is especially important because our data span a very long period and rely on data from a large number of countries. Over the period that we study, some cities changed their names (e.g., St. Petersburg became Leningrad), and a number of cities changed countries (e.g., Strasbourg was German in 1900 and 1914, and then became French). Furthermore, cities may be spelled in different languages in the faculty rosters and in papers in the Web of Science. For example, Rome is spelled using the German spelling “Rom” in *Minerva*, but it is spelled with either Italian (“Roma”), English (“Rome”), or German (“Rom”) spelling in the Web of Science, depending on the country of the journal.

We harmonize city and country information in multiple steps. The first step relies on the Google Maps API.

Step 1, part (i). We submit all city-country pairs (e.g., “London, United Kingdom”) that appear in the Web of Science to the API.⁵⁴ The Google Maps API returns a JSON file that contains the names of the city and country, the centroid coordinates for the city, and a location-type flag indicating the type of address found (e.g., “CITY” if the Google API found a city). Similarly, we geocode the faculty roster data with the Google Maps API. This also returns updated city and country names. Crucially, because we process addresses from both the Web of Science and the faculty rosters with the Google Maps API, we obtain harmonized addresses without spelling inconsistencies.

Step 1, part (ii). In some cases, the Google Maps API does not find the correct city and country. This usually occurs either because the name of a city or a country has changed over time (e.g., the name of Preßburg changed into Bratislava) or because of typos in the Web of Science data. We can identify such cases because the location type flag is “APPROXIMATE” instead of “CITY”. We improve the geocoding for these cases using the following 3-step procedure:

- (a) We structure the address before re-submitting it to the API (e.g., “’city’ : Preßburg, ’country’ : Hungary”).⁵⁵
- (b) For the cases without a result in step (a), we re-format the address-city string (e.g., “Preßburg,+Hungary”) and then re-submit it to the API.

⁵⁴The Web of Science contains not only the full affiliation address (e.g., “Loyola Univ Clinics, Mercy Hosp, Chicago, IL USA”) but also information on the city (“Chicago”) and country (“USA”) that the Web of Science team extracted from the affiliation address. We use the latter information for this step.

⁵⁵This option is not used as a baseline, since it reduces the match rate. Note: while today Bratislava is in Slovakia, it was part of the Austro-Hungarian Empire before WWI. Scientists from Bratislava therefore listed Preßburg, Hungary as their affiliation before WWI.

- (c) For the cases that do not return a result in steps (a) and (b), we re-submit the complete affiliation address (not just the city and country) from the Web of Science (e.g., “Loyola Univ Clinics, Mercy Hosp, Chicago, IL USA”) to the API.

Step 2. In some cases, the procedure above does not guarantee that the correct city and country have been found. We therefore rely on the Google Maps web interface, rather than the API, as a second step to improve the address data for addresses that appear misclassified. The advantage of the web interface, compared to the API, is that Google applies additional processing steps that improve the quality of the result.

To identify potentially misclassified addresses, we calculate the Levenshtein distance between the city name in the Web of Science and the city name returned by Google. If the Levenshtein distance is larger than three (i.e., more than three letters differ), we copy the full address from the Web of Science into the Google Maps web interface. If the web interface identifies the address, we extract the city and country from the website and use them as inputs for the Google Maps API (i.e., Step 1, part (i)). We further process the output from the Google Maps web interface with the Google Maps API because the web interface returns somewhat different city and country names than the API.

This processing of addresses ensures that addresses from the faculty rosters and the Web of Science are harmonized and that the records can be linked as described in Appendix A.4. below.

A.3.3. Predicting Academic Disciplines of Papers Using Paper Titles

To link papers from the Web of Science to the faculty rosters, we also match on discipline (see Appendix A.4. below). In preparation for this link, we therefore assign a discipline to each paper in the Web of Science.

The Web of Science assigns papers to academic fields (e.g., physics or general science) based on the journal they are published in, as opposed to assigning fields at the paper level. For 59% of the papers, this establishes a unique assignment to one of the disciplines in the faculty rosters (e.g., the journal *Acta Mathematica* is uniquely assigned to mathematics, which corresponds to a discipline in the faculty rosters). The remaining 41% of papers are published in journals that the Web of Science either assigns to multiple disciplines (e.g., the journal *Biometrika* is assigned to mathematics as well as biology) or to general science (e.g., *Nature*).⁵⁶

We thus train a multinomial logit classifier to assign a unique discipline to each paper, independently of where the paper was published. For example, the classifier assigns papers with a mathematics focus in *Biometrika* to mathematics, while assigning papers with a biology focus to biology.

We train the classifier using the words (unigrams), word pairs (bigrams), and word triplets (trigrams) from the titles of the 15,078,761 papers that the Web of Science assigns

⁵⁶In the Web of Science (and in the faculty rosters), statistics is a sub-discipline of mathematics.

to unique disciplines (e.g., the papers published in *Acta Mathematica*). In preparation for the training of the classifier, we remove very common words (stopwords) from the titles, as these contain little information. Next, we reduce words to their morphological roots using a stemmer. Afterwards, we transform the titles of each paper into a document 1,2,3-gram matrix \mathbf{X} of dimension $D \times V$, where D is the number of papers in our data and V is the total number of unique unigrams, bigrams, and trigrams in all titles:

$$\mathbf{X} \equiv \text{document 1,2,3-gram matrix} = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,V} \\ w_{2,1} & \ddots & & w_{2,V} \\ \vdots & & \ddots & \vdots \\ w_{D,1} & w_{D,2} & \cdots & w_{D,V} \end{pmatrix}.$$

The individual entries $w_{d,v}$ represent the number of times n-gram v appears in document d . The individual entries in the matrix are then reweighted by their term-frequency-inverse-document-frequency (tf-idf) such that $\text{tf-idf}(w_{d,v}) = (1 + \log(w_{d,v})) \times \left(\log \left(\frac{1+D}{1+d_v} \right) + 1 \right)$, where d_v is the number of documents n-gram v appears in at least once. This reweighting reduces the weights of n-grams that appear in many titles of papers (e.g., method).

The multinomial logit classifier then learns to predict disciplines based on the 1,2,3-gram matrix \mathbf{X} , where the dependent variable y_d is the discipline of the paper. To avoid overfitting, we include L2 regularization in the classifier. As is standard in the machine-learning literature, we choose the optimal regularization strength using 10-fold cross-validation, evaluated on the basis of the F1-score.⁵⁷ The final classifier achieves a within-sample F1-score of 0.99 and an out-of-sample F1-score of 0.81. As some biochemistry papers are published in chemistry journals, we would expect an F1-score of less than one. Using our classifier, we predict a unique discipline for the 10,508,299 papers that the Web of Science originally assigns to multiple disciplines (on the basis of the journal).

A.4. Merging Academics to Web of Science

We match papers from the Web of Science to the faculty rosters using a nine-step procedure. As mentioned in the main text, we match papers from the Web of Science within a \pm five-year window around the year of the corresponding cohort of academics. For example, for scientists listed in the 1914 cohort, we only match papers published between 1909 and 1919. Within these windows, we match the Web of Science data to each cohort of academics using the following sequential procedure:

1. Merge using: i) full surname, ii) full first name, iii) discipline, iv) country, v) city.
2. Merge using: i) full surname, ii) all initials, iii) discipline, iv) country, v) city.

⁵⁷The F1-score is defined as $F1 = \frac{TP}{TP + 0.5(FP + FN)}$, where TP is the number of true positives, FP the number of false positives, and FN the number of false negatives. To speed up the training process, the 10-fold cross-validation is run on a 20% random subset of the data before training the final classifier on the full data.

3. Merge using: i) full surname, ii) first initial, iii) discipline, iv) country, v) city. Scientists and journals do not publish a consistent number of initials. We therefore exclude matches in which the initials indicate that the paper in the Web of Science was not published by the scientist listed in the faculty rosters. We use the following rule to exclude incorrect matches: Denote the string of initials of a scientist in the faculty rosters by s and the string of initials of the scientist in the Web of Science by p :

- (a) If the number of initials in s and p is identical ($|s| = |p|$), but the initials differ ($s \neq p$) we exclude the match. For example, a match of a scientist listed in the faculty rosters with initials “A.A.” will not be merged to a paper published by someone with initials “A.B.” (Note: as described in step 3, we only consider matches where the full surname, subject, country, and city match.)
- (b) If the number of initials in s and p is not identical ($|s| \neq |p|$), we exclude matches in which not all letters from the shorter set of initials appear in the other in the same order. To implement this rule, we compute the Levenshtein distance between the two strings of initials s and p ($lev(s, p)$). If $lev(s, p)$ is larger than the difference in the length of the strings, i.e., $lev(s, p) > ||s| - |p||$, the match is excluded. For example, a scientist listed in the faculty rosters with initials “A.B.” will not be merged to a paper published by someone with initials “A.C.D.” or “A.C.B.”, but it will be merged to someone with initials “A.B.C.”

4.-6. We then repeat steps 1-3, but remove the city from the merge conditions.

7.-9. We repeat steps 1-3, but additionally remove the country from the merge conditions.

If one of the authors of a paper is matched to a scientist in an earlier (and thus more restrictive) step, this particular author will no longer be considered in any following step. We account for the fact that some papers are merged to multiple scientists by weighting the papers by the total number of matches.

For the period covered by our paper, the Web of Science rarely provides a unique assignment of the addresses reported on a paper to its coauthors. For example, if a paper has two coauthors and they are affiliated with different institutions, the Web of Science usually does not specify which coauthor is affiliated with which institution. We therefore assign each address reported on a paper to all coauthors of the paper. If more than one address is associated with a paper, we perform a many-to-many merge between addresses and coauthors. As we show in Table 3, the results are robust to considering only scientists with a unique surname, first initial, and discipline across all universities in the world.

A.5. Number of Universities in Minerva and WHED

Table A.3: Universities in Minerva and WHED

Country	Number of Universities		Country	Number of Universities	
	Minerva	WHED		Minerva	WHED
Afghanistan	2	3	Libya	8	2
Albania	4	5	Lithuania	9	9
Algeria	11	5	Luxembourg	3	1
Angola	4	0	Madagascar	3	1
Argentina	49	23	Malawi	1	1
Armenia	0	13	Malaysia	7	5
Australia	35	29	Mali	1	1
Austria	40	22	Malta	1	1
Azerbaijan	1	16	Martinique	2	1
Bangladesh	101	8	Mauritius	2	0
Barbados	2	0	Mexico	69	101
Belarus	1	25	Moldova	3	7
Belgium	58	25	Mongolia	1	8
Benin	1	0	Montenegro	1	0
Bolivia	9	9	Morocco	10	6
Bosnia and Herzegovina	1	1	Mozambique	1	0
Brazil	150	104	Myanmar (Burma)	10	12
Brunei	1	0	Nepal	21	1
Bulgaria	22	24	Netherlands	58	19
Burundi	2	0	New Zealand	12	6
Cambodia	1	4	Nicaragua	5	4
Cameroon	4	0	Nigeria	16	4
Canada	133	57	North Korea	1	33
Chad	2	0	North Macedonia	1	1
Chile	23	15	Norway	16	12
China	27	561	Pakistan	150	24
Colombia	35	49	Palestine	0	3
Costa Rica	2	3	Panama	2	2
Croatia	14	1	Papua New Guinea	1	0
Cuba	9	6	Paraguay	7	2
Czechia	47	20	Peru	32	18
Ivory Coast	3	3	Philippines	46	436
Democratic Republic of the Congo	7	7	Poland	80	77
Denmark	13	12	Portugal	33	9
Dominican Republic	3	3	Puerto Rico	4	10
Ecuador	11	11	Republic of the Congo	2	0
Egypt	27	7	Romania	70	35
El Salvador	3	2	Russia	90	340
Estonia	3	5	Rwanda	1	0
Ethiopia	2	4	Samoa	4	0
Fiji	1	1	Saudi Arabia	4	2
Finland	16	8	Senegal	2	1
France	351	145	Serbia	14	2
French Guyana	0	1	Sierra Leone	6	0
Gabon	0	1	Singapore	6	3
Georgia	1	12	Slovakia	12	13
Germany	281	133	Slovenia	7	1
Ghana	6	10	Solomon Islands	1	0
Greece	14	12	Somalia	1	1
Guam	1	1	South Africa	28	12
Guatemala	3	1	South Korea	43	96
Guinea	1	0	Spain	137	39
Guyana	3	0	Sri Lanka	87	5
Haiti	3	3	Sudan	7	6
Honduras	2	3	Sweden	27	15
Hong Kong	3	4	Switzerland	33	14
Hungary	64	24	Syria	6	2
Iceland	6	3	Taiwan	24	42
India	1971	103	Tajikistan	0	8
Indonesia	46	62	Tanzania	6	2
Iran	17	23	Thailand	17	43
Iraq	19	1	Trinidad and Tobago	2	2
Ireland	22	8	Tunisia	13	2
Israel	25	21	Turkey	20	16
Italy and Vatican City	229	61	Turkmenistan	0	4
Jamaica	2	4	USA	1540	1554
Japan	291	311	Uganda	3	3
Jordan	7	0	Ukraine	23	147
Kazakhstan	0	26	United Kingdom	309	111
Kenya	4	10	Uruguay	3	7
Kuwait	1	1	Uzbekistan	0	24
Kyrgyzstan	0	10	Venezuela	20	11
Latvia	3	11	Vietnam	11	28
Lebanon	11	10	Zambia	3	1
Lesotho	2	1	Zimbabwe	6	2
Liberia	3	2	Total	7477	5520

Notes: The Table shows the number of universities in Minerva in all cohorts from 1900 until 1969. It compares Minerva to the World Higher Education Database (WHED), available at <https://www.whed.net/home.php>. Minerva starts listing universities around 5-15 years after they are founded. The last Minerva cohort was published between 1966 and 1969. To account for this delay, we count universities in WHED if they were founded before 1961. The WHED does not include micro-data on individual academics.

A.6. Benchmarking the Minerva Data

To the best of our knowledge, there are no comparable data covering academics worldwide over many decades. To provide evidence on the coverage of *Minerva*, we benchmark the data in two ways. First, we show that the number of universities covered in *Minerva* is similar to the number of universities included in the World Higher Education Database, which was designed as a census of all worldwide universities (WHED (2024); see Appendix Table A.3). This comparison indicates that *Minerva* captures a very large fraction of worldwide universities. For most countries, *Minerva* covers more universities than are listed in the World Higher Education Database. Coverage is lower for Russia and other countries that were part of the Soviet Union because the Soviet Union stopped reporting academics to *Minerva* from the 1938 cohort onwards. Coverage is also lower for China, which established a modern university sector only during the 20th century.

As the WHED does not include microdata on individual academics, we perform additional benchmarking exercises on smaller datasets that list individual academics in some universities and time periods.

A.6.1. Benchmarking Against Rossiter (1982)/American Men of Science (1938)

Rossiter (1982), p. 182 reports female scientists in twenty major U.S. universities for the year 1938. The data are based on women listed in the historical publication *American Men of Science* (AMS), 6th edition, 1938. The data contain all female scientists who are listed in the AMS for twenty leading U.S. institutions.

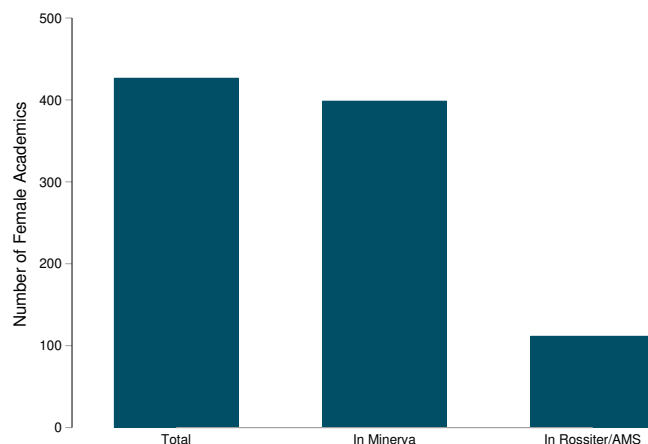
For the benchmarking exercise, we extract all female scientists who are at least assistant professors who are listed in these twenty universities in *Minerva* 1938. We then cross-check all names and identify women listed in both sources. Both sources combined list a total of 427 different female academics, which we take as the best available information for the total number of women in these twenty universities in 1938 (first bar, Appendix Figure A.2). Of these, 399 (93%) are listed in *Minerva* (second bar).⁵⁸ In contrast, Rossiter, on the basis of the *American Men of Science*, only lists 112 (26%) of them (third bar). This indicates that *Minerva* 1938 has a much more comprehensive coverage of academics in the top twenty U.S. universities for 1938 than the *American Men of Science*.

A.6.2. Benchmarking Against German University Catalog Data

We also benchmark the *Minerva* data against data from semi-official German university calendars listing all academics who were lecturing at any German university during the winter semester 1937/38. The university calendar was published by J.A. Barth, who collected official university calendars from all 32 German universities and compiled them into one volume called *Kalender der reichsdeutschen Universitäten und Hochschulen*.

⁵⁸The 7% missing female academics in *Minerva* are due to the following reasons: 1) in 1938 Minnesota (one of the 20 universities) only reported full professors in *Minerva* but Rossiter reports 9 female assistant or associate professors for Minnesota. 2) even though both sources were published in 1938 they may report faculty based on slightly different cutoff dates.

Figure A.2: Benchmarking *Minerva* Against Rossiter (1982) / American Men of Science (1938)

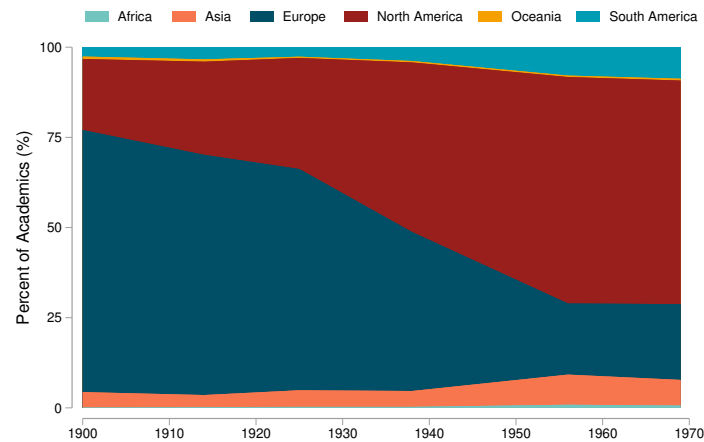


Notes: The Figure shows the number of female scientists in twenty major U.S. universities for the year 1938 and how they are covered by different sources.

We extract all physicists, chemists, and mathematicians in the same way as Waldinger (2012). Overall, these data contain 866 scientists in the three fields for the winter semester 1937/38. We then match these scientists to *Minerva*, matching on the surname, first name, discipline, and university. Of the 866 scientists, we can match 853 in *Minerva*, a match rate of 98.5%, suggesting that the coverage of *Minerva* was very comprehensive.

B Further Results

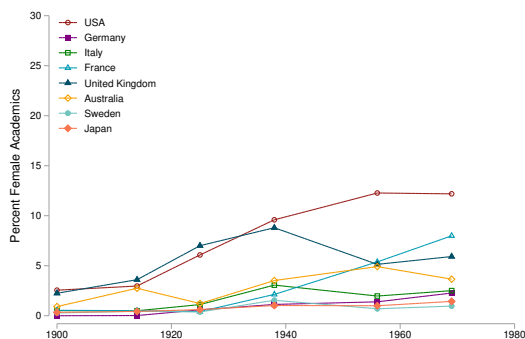
Figure B.1: Percent of Academics by Continent over Time, 1900-1969



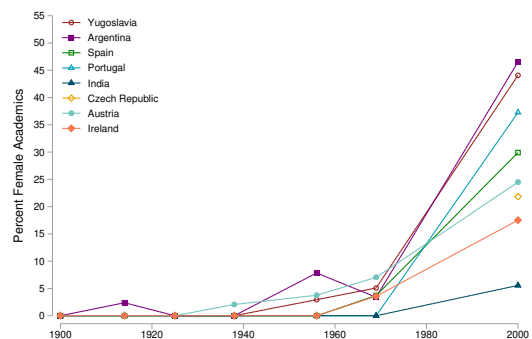
Notes: The Figure shows the percentage of academics by continent across all universities and disciplines for the period 1900-1969. The data were collected by the authors from various volumes of *Minerva*, see Section 1 for details.

Figure B.2: Percent of Female Academics by Country over Time, Additional Evidence

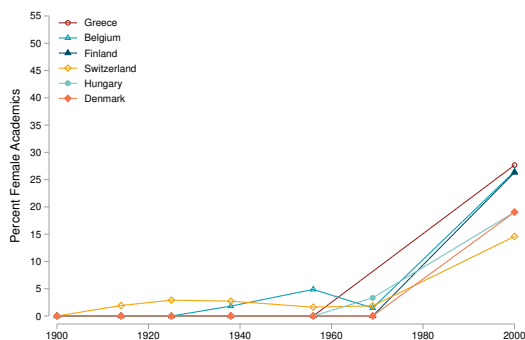
(a) Excluding Women's Colleges



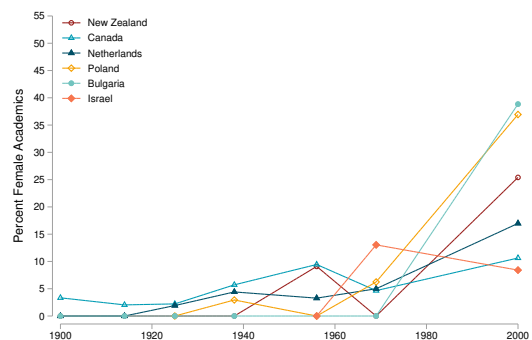
(b) Additional Countries



(c) Additional Countries

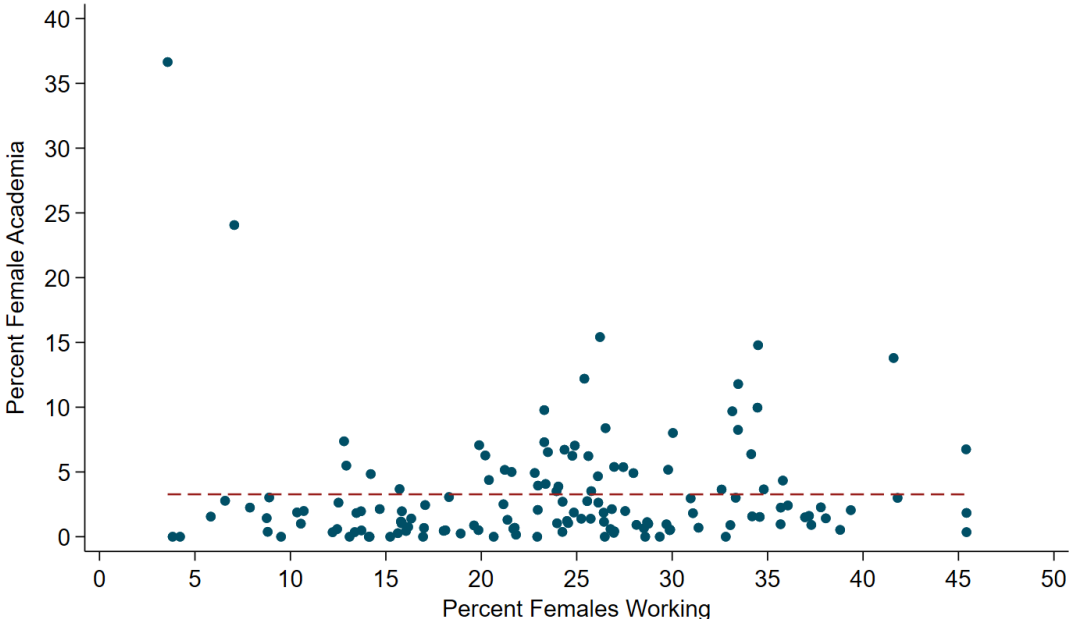


(d) Additional Countries



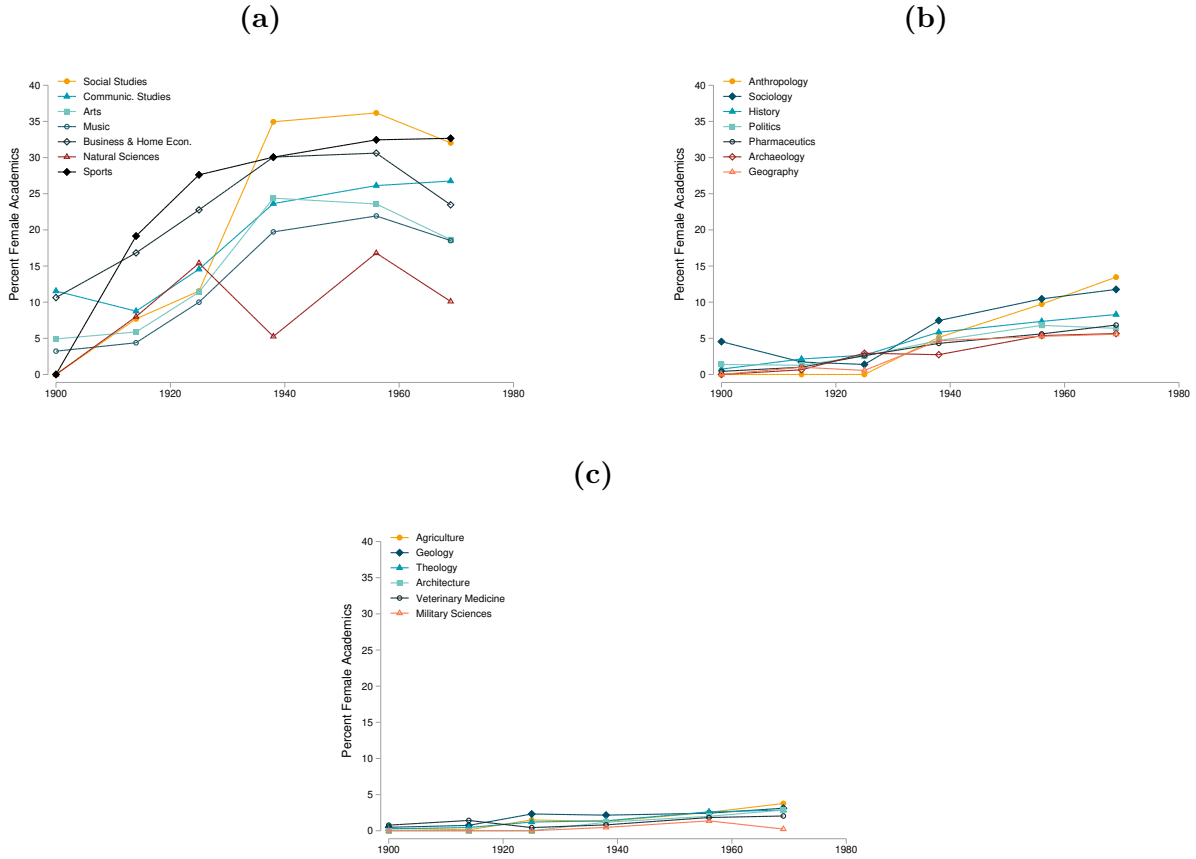
Notes: The Figure shows the percentage of female academics by country over time. Panel (a) plots the percentage of female academics across all universities and disciplines until 1969 (Sample 1), excluding women's colleges. Panels (b)-(d) plot the percentage of female scientists in prestigious universities (Sample 3).

Figure B.3: Correlation of Percent Female Academics with General Labor Force Participation of Women



Notes: The Figure shows the percentage of female academics (vertical axis) and the percentage of women employed in the labor market (horizontal axis). Each dot represents a country-cohort pair (e.g., USA in 1969). The data were collected by the authors from various volumes of *Minerva* and department websites, see Section 1 for details. The data on general labor force participation come from *International Historical Statistics: Africa and Asia, The Americas and Australasia, and Europe* (Mitchell (1982, 1983, 1993)). The female share is calculated as the number of women working in all sectors among all women aged 15 and above. The data for the United States come from Killingsworth and Heckman (1986), Table 2.1.

Figure B.4: Percent of Female Academics, Additional Disciplines



Notes: The Figure shows the percentage of female academics across additional disciplines not reported in the main paper for the period 1900-1969 (Sample 1). Note that the figure combines Business and Home Economics into one line. The data were collected by the authors from various volumes of *Minerva* and department websites, see Section 1 for details.

C Linking Gender Gaps in Hiring and in Publications

In this Appendix, we show that regression (3) in the main paper can be micro-founded with a model along the lines of Roy (1951).

C.1. A Roy Model

The model allows for (i) selection on unobservables in the hiring market, (ii) gender bias in hiring, and (iii) gender bias in the publication market. These factors contribute to the gender gap in publications through (a) *indirect* effects of selection and gender bias in the hiring market, and (b) *direct* effects of gender bias in the publication market.⁵⁹

⁵⁹A more general model also incorporating gender gaps in citations (Section 4) and promotions (Section 5) would have to impose too many (and controversial) assumptions to remain tractable. For example, the effect of selection in publishing on citations is unclear from a modeling perspective. It could be positive for some women and negative for others. On the one hand, women published fewer papers and, hence, some women may have published only their highest-quality ideas. On the other hand, women may have been impeded from carrying out high-quality research, e.g., because of a disproportionate share of housework, which may have resulted in fewer and lower-quality publications. Combining these two countervailing forces in a model would have to rely on unreasonable assumptions.

Selection in the Hiring Market

At the hiring stage, denoted by 0, any academic position i can be filled with a woman W or a man M . Women and men face differential barriers until they are hired as academics. Such barriers may be institutional (e.g., certain high schools, many undergraduate programs, and most PhD programs did not accept women for a large part of the 20th century, Rossiter 1982) or stem from gender differences in exposure to academic role models (e.g., Bell et al. 2019). We refer to any such gender bias in hiring as Δ_0 . The latent value of hiring a woman is:

$$Y_{0i}^W = r_W(X_i^W) + \epsilon_{0i}^W, \quad (\text{C.1})$$

while that of hiring a man is:

$$Y_{0i}^M = r_M(X_i^M) + \Delta_0 + \epsilon_{0i}^M, \quad (\text{C.2})$$

where $r_g(\cdot)$ is an unknown function of observable characteristics X_i^g , $g \in \{W, M\}$, Δ_0 is a possible gender bias in hiring, and ϵ_{0i}^g is the unobserved component of these latent valuations. As a result, academic position i is filled with a woman if (C.1) is greater than (C.2):

$$\begin{aligned} Y_{0i} &= (\epsilon_{0i}^W - \epsilon_{0i}^M) + r_W(X_i^W) - r_M(X_i^M) - \Delta_0 \\ &= \epsilon_{0i} - r_0(X_i) > 0, \end{aligned} \quad (\text{C.3})$$

with $\epsilon_{0i} \equiv \epsilon_{0i}^W - \epsilon_{0i}^M$, $X_i \equiv (X_i^W, X_i^M)$, and $r_0(X_i) \equiv r_M(X_i^M) - r_W(X_i^W) + \Delta_0$, so that, keeping everything else fixed, when $\Delta_0 > 0$ women need to overcome the additional hurdle, or gender bias, Δ_0 to be hired. The function $r_0(\cdot)$ can be nonparametric with respect to X_i .⁶⁰ We assume that $r_0(\cdot)$ is not itself a function of the share of women among all hired scientists (s_0^W). The error term ϵ_{0i} is distributed according to F , an unknown c.d.f. assumed to be invertible. As a result, the share of women among all hired scientists s_0^W , conditional on observables, is:

$$\begin{aligned} s_0^W(X_i) &= \Pr[Y_{0i} > 0 \mid X_i] = \Pr[\epsilon_{0i} > r_0(X_i) \mid X_i] = 1 - \Pr[\epsilon_{0i} \leq r_0(X_i) \mid X_i] \\ &= 1 - F(r_0(X_i)). \end{aligned} \quad (\text{C.4})$$

Observable characteristics X_i could capture SAT scores, college GPA, specialization of undergraduate degree, or differential treatment of boys and girls during childhood. Such data are not available at a worldwide scale over the 20th century and, if available, they would be affected by various sources of selectivity and measurement error. As a consequence, we assume $X_i = X_0$ for all i 's, with X_0 some constant value such that $r_M(X_0^M) = r_W(X_0^W)$.⁶¹

⁶⁰For example, $r_0(\cdot)$ trivially fits the linear specification $r_0(X_i) = X_i\beta_0 + \Delta_0$, but can also take the more general form $r_0(X_i) = h(X_i, \beta_0) + \Delta_0$, with $h(X_i, \beta_0)$ any function of X_i and the parameter β_0 (of any dimension).

⁶¹For example, if $r_M(\cdot) = r_W(\cdot)$, this would hold for any X_0 such that $X_0^M = X_0^W$.

It then follows that equation (C.4) simplifies to $s_0^W = 1 - F(\Delta_0)$. The invertibility of F then implies that $\Delta_0 = F^{-1}(1 - s_0^W)$, a fact we use below in equations (C.6) and (C.7).

Publication Market

Conditional on academic position i being filled by either a woman or a man at the hiring stage (stage 0), we observe publication outcomes for all hired scientists (stage 1). We refer to any gender bias in publications as Δ_1 . The publication outcomes for women and men are:

$$\begin{aligned} Y_{1i}^W &= Z_i^W \beta_1 + \epsilon_{1i}^W && \text{if } Y_{0i} > 0 \\ Y_{1i}^M &= Z_i^M \beta_1 + \Delta_1(s_0^W) + \epsilon_{1i}^M && \text{if } Y_{0i} \leq 0, \end{aligned} \tag{C.5}$$

where Z_i^g , $g \in \{W, M\}$, are observable characteristics and ϵ_{1i}^g is the unobserved component of the publication outcome Y_{1i}^g . If academic position i is filled by a woman, we observe the publication outcome of a woman, otherwise we observe the outcome of a man. Since for any i we cannot observe the counterfactual publication outcome (i.e., the publications if position i had been filled by the other gender), equation (C.5) will be subject to selection on unobservables if the error terms in (C.3) and (C.5) are correlated.

$\Delta_1(s_0^W)$ is the gender bias in publications, which may reflect gender imbalances in working conditions, preferences, or other constraints that differentially affected publications, such as discrimination in the peer-review process or in the workplace, or differences in childcare responsibilities. As the notation $\Delta_1(s_0^W)$ highlights, we allow the gender bias in publications to be a function of the share of female scientists. The gender bias in publications can be a function of the share of women because of external forces (e.g., social norms) that are making it easier for women to both get hired and publish in academia or because the hiring of women itself improves their publishing opportunities (e.g., by increasing the presence of women among editors and referees). In our estimation, we will capture the combined effect of these two forces.

Publications Conditional on Gender

Because selection at the hiring stage determines which individuals are subsequently observed at the publication stage, the conditional expectation of Y_{1i}^W given $Y_{0i} > 0$ is:

$$\begin{aligned} \mathbb{E}[Y_{1i}^W | Z_i^W, Y_{0i} > 0] &= Z_i^W \beta_1 + \mathbb{E}[\epsilon_{1i}^W | \epsilon_{0i} > \Delta_0] \\ &= Z_i^W \beta_1 + \tilde{g}_W(\Delta_0) = Z_i^W \beta_1 + \tilde{g}_W(F^{-1}(1 - s_0^W)) \\ &= Z_i^W \beta_1 + g_W(s_0^W), \end{aligned} \tag{C.6}$$

where $\tilde{g}_W(\Delta_0) \equiv \mathbb{E}[\epsilon_{1i}^W | \epsilon_{0i} > \Delta_0]$ and $g_W(s_0^W) \equiv \tilde{g}_W(F^{-1}(1 - s_0^W))$. Analogously, the conditional expectation of Y_{1i}^M is:

$$\begin{aligned}
\mathbb{E}[Y_{1i}^M | Z_i^M, Y_{0i} \leq 0] &= Z_i^M \beta_1 + \Delta_1(s_0^W) + \mathbb{E}[\epsilon_{1i}^M | \epsilon_{0i} \leq \Delta_0] \\
&= Z_i^M \beta_1 + \Delta_1(s_0^W) + \tilde{g}_M(\Delta_0) \\
&= Z_i^M \beta_1 + \Delta_1(s_0^W) + \tilde{g}_M(F^{-1}(1 - s_0^W)) \quad (\text{C.7}) \\
&= Z_i^M \beta_1 + \Delta_1(s_0^W) + g_M(s_0^W) \\
&= Z_i^M \beta_1 + G_M(s_0^W),
\end{aligned}$$

where $\tilde{g}_M(\Delta_0) \equiv \mathbb{E}[\epsilon_{1i}^M | \epsilon_{0i} \leq \Delta_0]$ and $g_M(s_0^W) \equiv \tilde{g}_M(F^{-1}(1 - s_0^W))$. As a result, the expectation of the observed publication outcome can be obtained by combining equations (C.6) and (C.7):

$$\mathbb{E}[Y_{1i}^M | Z_i, Y_{0i} \leq 0] + \text{Female}_i \times \left[\mathbb{E}[Y_{1i}^W | Z_i, Y_{0i} > 0] - \mathbb{E}[Y_{1i}^M | Z_i, Y_{0i} \leq 0] \right], \quad (\text{C.8})$$

where Female_i is an indicator that denotes whether scientist i is a woman. If a man is hired (i.e., $\text{Female}_i = 0$), the observed publication outcome is $\mathbb{E}[Y_{1i}^M | Z_i, Y_{0i} \leq 0]$. Alternatively, if a woman is hired (i.e., $\text{Female}_i = 1$), the observed publication outcome is $\mathbb{E}[Y_{1i}^W | Z_i, Y_{0i} > 0]$ because the two expectations for men $\mathbb{E}[Y_{1i}^M | Z_i, Y_{0i} \leq 0]$ cancel out.

Under the assumption that observed position characteristics, such as cohort, country, and discipline, affect publication outcomes identically for men and women (i.e., $Z_i^W = Z_i^M = Z_i$), the expression above simplifies to:

$$Z_i \beta_1 + G_M(s_0^W) + \text{Female}_i \left[g_W(s_0^W) - G_M(s_0^W) \right]. \quad (\text{C.9})$$

We control for $Z_i \beta_1$ using cohort, country, and discipline, or university fixed effects. Without further assumptions, it is not possible to separately identify the various components of $g_W(\cdot)$ and $G_M(\cdot)$. To avoid unnecessarily strong functional form restrictions, we directly approximate these functions by polynomials of degree \mathcal{K} in the share of female scientists: $g_W(s_0^W) = \sum_{\kappa=0}^{\mathcal{K}} \theta_{\kappa}^W \times (s_0^W)^{\kappa}$ and $G_M(s_0^W) = \sum_{\kappa=0}^{\mathcal{K}} \theta_{\kappa}^M \times (s_0^W)^{\kappa}$. This yields the following expression:

$$Z_i \beta_1 + \sum_{\kappa=0}^{\mathcal{K}} \theta_{\kappa}^M \times (s_0^W)^{\kappa} + \text{Female}_i \times \left[\sum_{\kappa=0}^{\mathcal{K}} (\theta_{\kappa}^W - \theta_{\kappa}^M) \times (s_0^W)^{\kappa} \right]. \quad (\text{C.10})$$

Estimating Regression

We estimate equation (C.10) using a quadratic polynomial ($\mathcal{K} = 2$) and by allowing the

gender gap function to vary by period p , for $p = 1$ (1900-38), 2 (1956-69), 3 (2000):⁶²

$$\begin{aligned}
\text{Pub}_{it} = & \delta + \text{Female}_i \times 1 [t(i) = 1900/38] \times \underbrace{\left[\gamma_{01} + \gamma_{11}s_{0\ell(i)}^W + \gamma_{21} (s_{0\ell(i)}^W)^2 \right]}_{g_1(s_{0\ell(i)}^W) \equiv \text{gender gap function in 1900-38}} \\
& + \text{Female}_i \times 1 [t(i) = 1956/69] \times \underbrace{\left[\gamma_{02} + \gamma_{12}s_{0\ell(i)}^W + \gamma_{22} (s_{0\ell(i)}^W)^2 \right]}_{g_2(s_{0\ell(i)}^W) \equiv \text{gender gap function in 1956-69}} \\
& + \text{Female}_i \times 1 [t(i) = 2000] \times \underbrace{\left[\gamma_{03} + \gamma_{13}s_{0\ell(i)}^W + \gamma_{23} (s_{0\ell(i)}^W)^2 \right]}_{g_3(s_{0\ell(i)}^W) \equiv \text{gender gap function in 2000}} \\
& + \text{Experience}_{it} \boldsymbol{\delta}_{\text{exp}} + \text{FE}(i, t) + \varepsilon_{it},
\end{aligned} \tag{C.11}$$

where Pub_{it} measures the standardized number of papers published by scientist i in cohort $t(i)$, $s_{0\ell(i)}^W$ is the share of female scientists in i 's country-cohort pair $\ell(i)$ (e.g., the United States in 2000), and each $\gamma_{\kappa p}$ corresponds to the gender gap $\theta_{\kappa p}^W - \theta_{\kappa p}^M$, $\kappa = 0, 1, 2$, which we estimate separately for each period $p = 1, 2, 3$.⁶³ We control for the term $Z_i\beta_1 + \sum_{\kappa=0}^2 \theta_{\kappa}^M \times (s_0^W)^{\kappa}$ in equation (C.10) with cohort-discipline-country (or finer) fixed effects (see definition (2) in the main text for details) and experience-by-discipline indicators.⁶⁴ While the fixed effects in regression (C.11) capture the average number of publications among male scientists, the gender gap functions capture any systematic difference in the publications of women versus men.

Importantly, regression (C.11) does not “force” any specific shape for $g_p(s_{0\ell}^W)$ on the data, in that a lack of gender gaps, an increasing or decreasing relationship, a straight line, a U, or an inverted U could all be estimated. The estimated parameters $(\hat{\gamma}_{0p}, \hat{\gamma}_{1p}, \hat{\gamma}_{2p}, p = 1, 2, 3)$ determine the shape of the gender gap as a function of the share of female scientists.

C.2. Estimation Results: “Gender U”

We report estimates of regression (C.11) for Sample 3 in Appendix Table C.1. Estimating the gender gap function is empirically demanding given the extensive set of fixed effects included in all specifications. Even the baseline specification in column 1 is stringent, as it absorbs cohort-discipline-country fixed effects (e.g., physics in the United States in 2000). Despite this stringent saturation, all six coefficients that govern the shape of the gender

⁶²We obtain qualitatively similar estimates of the gender gap function by approximating it with higher-degree polynomials, such as $\mathcal{K} = 3$ or $\mathcal{K} = 4$.

⁶³To reduce measurement error in the regressors $s_{0\ell}^W$ and $(s_{0\ell}^W)^2$ due to very low numbers of female scientists in some cohorts, disciplines, and countries, we compute the share of female scientists $s_{0\ell}^W$ at the level of the country-cohort pair ℓ , rather than at the finer level of the cohort-discipline-country.

⁶⁴This assumes that Z_i varies at most at the level of the fixed effects included in regression (C.11). In addition, because we measure s_0^W at the country-cohort pair level, the fixed effects always fully control for the occurrence of $\sum_{\kappa=0}^2 \theta_{\kappa}^M \times (s_0^W)^{\kappa}$ not interacted with Female_i in equation (C.10).

gap function— γ_{1p} and γ_{2p} for $p = 1, 2, 3$ —are statistically significant at conventional levels.

The specifications in columns 2 and 3 impose progressively more stringent controls. Column 2 adds university-by-cohort fixed effects (e.g., Harvard in 2000), while column 3 adds department-by-cohort fixed effects (e.g., physics at Harvard in 2000). These specifications are particularly demanding given the limited identifying variation in earlier cohorts. Prior to 2000, relatively few departments had hired any women, thus identification—especially within universities and even more so within departments—relies on a restricted subset of units. Departments that had not hired at least one woman do not contribute to within-unit identification, further reducing effective variation. As a consequence, statistical precision declines in earlier periods under the most saturated specifications.

Importantly, however, the point estimates are remarkably stable across columns. In the 2000 cohort, female hiring had become more widespread across departments and universities, providing sufficient within-department variation. For this cohort, the coefficients that govern the shape of the gender gap function, γ_{13} and γ_{23} , remain precisely estimated and statistically significant.

To interpret the results in Appendix Table C.1, note that the estimates $\hat{\gamma}_{0p}$ do *not* represent the total gender bias in publications. In this model, the gender gap function is given by: $\hat{g}_p(s_{0\ell}^W) = \hat{\gamma}_{0p} + \hat{\gamma}_{1p}s_{0\ell}^W + \hat{\gamma}_{2p}(s_{0\ell}^W)^2$.

To assess how the gender gap in publications varies with the share of female scientists, we plot the estimated gender gap function as calculated from column 1 of Appendix Table C.1 against observed values of $s_{0\ell}^W$ in Sample 3.⁶⁵ Each dot in the figure represents the predicted gender gap in publications for period p as a function of the share of female scientists in each cohort-country pair (e.g., the United States in 2000). The figure suggests a U-shaped relationship between the gender gap in publications and the share of female scientists. As discussed in the main text, we refer to this relationship as the “gender U.” The gender U arises because the estimates $\hat{\gamma}_{1p}$ are negative and the estimates $\hat{\gamma}_{2p}$ are positive for each p .

In interpreting the gender U, it is crucial to recognize that publications are an imperfect proxy for true research ability. Observed publications therefore reflect not only individual productivity but also a range of institutional and structural factors, including discrimination in the peer-review process, unequal access to key inputs such as laboratory space and research funding, and asymmetries in coauthorship networks. These constraints likely varied substantially across countries and plausibly evolved in tandem with the expansion of women’s participation in academia.

The estimates $\hat{\gamma}_{1p}$ and $\hat{\gamma}_{2p}$ could be interpreted as capturing selection in true research ability alone, but only under the restrictive assumption that such factors did not change systematically with the female share in academia. A more plausible view is that these institutional features shifted as women’s representation increased during the 20th

⁶⁵Panel (b) of Figure 8 plots a similar relationship but for Sample 2.

Table C.1: Individual-Level Publication Gaps and the Share of Females

	(1)	(2)	(3)
Dependent Variable	Standardized Publications		
Female (1900/38)			
γ_{01}	0.007 (0.160)	-0.015 (0.183)	-0.001 (0.214)
γ_{11}	-12.593** (5.386)	-9.365 (6.093)	-11.280 (7.034)
γ_{21}	90.248* (50.163)	67.432 (52.119)	81.744 (56.546)
Female (1956/69)			
γ_{02}	-0.037 (0.179)	-0.017 (0.175)	-0.147 (0.172)
γ_{12}	-10.875** (4.915)	-8.711* (4.876)	-5.012 (4.781)
γ_{22}	61.310** (28.618)	44.499 (27.822)	24.588 (27.323)
Female (2000)			
γ_{03}	0.219 (0.134)	0.192 (0.121)	0.176 (0.110)
γ_{13}	-3.342*** (0.958)	-2.941*** (0.884)	-2.647*** (0.811)
γ_{23}	5.161*** (1.602)	4.364*** (1.510)	3.757*** (1.417)
Observations	88,809	88,809	88,809
R-squared	0.018	0.057	0.112
Experience×Discipline	Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes	
Cohort×University FE		Yes	
Cohort×Discipline×University FE			Yes

Notes: The Table shows estimation results for regression (C.11). The sample includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in prestigious universities until 2000 (Sample 3). The dependent variable equals the standardized number of publications in a \pm five-year window around a cohort (i.e., 1995-2005 for a scientist listed in 2000). The main explanatory variables are indicators that equal 1 if the scientist is a woman, interacted with second-degree polynomials of $s_{0\ell}^W$. All regressions also control for different sets of fixed effects (see definition (2) in the main text for details) and experience-by-discipline indicators. Standard errors are clustered at the discipline-country level, with 184 clusters. Significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

century. In that case, the estimated coefficients reflect not only selection in true research ability, but also contemporaneous changes in discrimination, access to resources, and other determinants of publication opportunities.

Accordingly, the gender U should not be interpreted as arising solely from compositional changes in the ability distribution of men and women. Rather, it likely reflects the joint influence of selection and evolving institutional constraints. The relative importance of these channels determines the observed shape of the gender gap over time.

For example, the downward-sloping part of the gender U could have arisen due to positive selection of women. Alternatively, a higher female share may have been associated with *more* discrimination against women in the publication market. Both of these channels would have increased the publication gap as a function of the female share; however, the second channel seems less plausible. Similarly, the upward-sloping part of the gender U could have arisen due to negative selection of men or a rising empowerment of women. A rising empowerment of women could have reduced discrimination in the publication

market and the allocation of resources, or improved coauthoring opportunities with other women. All of these channels could have contributed to a closing publication gap as a function of the share of female scientists. As outlined in the main text, a plausible interpretation for the gender U is that the most talented women entered academia first (*selection effect*), and a greater representation of women in academia was accompanied by more publishing opportunities for women (*empowerment effect*).

The alternative explanation that the gender gap in publications widened at first due to an increase in discrimination against women in the publication market and then closed again due to negative selection of men strikes us as less plausible for at least two reasons. First, with rising shares of female scientists, increasing discrimination against women in the publication market is inconsistent with the evidence in Card et al. (2022) and Card et al. (2023), who find increasing *recognition* of female academics over time as measured by elections to prestigious scientific societies. Second, negative selection of men would imply that the best male scientists left academia in response to female entry. However, panel (a) of Figure 9 in the main text shows that men’s average publications did not change with the share of female scientists, suggesting that selection of men into academia was unaffected by female entry.

D Further Details and Results: Citation Gaps

D.1. Further Details: Predicted Citations

As outlined in the main text, we aim to account flexibly for the topic of each paper, which could influence citations. We estimate a ridge regression that uses the words (unigrams) and word pairs (bigrams) that appear in the title of the 749,465 scientific papers, which we match to at least one scientist in our data. The model learns how many citations, on average, papers in finely-grained research fields typically receive.⁶⁶

In preparation for the ridge regression, we remove stopwords from the titles and reduce all words to their morphological roots using a stemmer. We then transform the titles of each paper into a document 1,2-gram matrix \mathbf{X} of dimension $D \times V$, where D is the number of papers and V is the total number of unique unigrams and bigrams plus 30 indicators for the length of titles.⁶⁷

The model minimizes equation (4) to identify the n-grams that have the highest predictive power for citations. The regularization term λ stabilizes the estimates and prevents overfitting in this high-dimensional setting, where the number of n-gram regressors is large relative to the number of papers in each discipline-cohort cell. We choose the optimal regularization strength using 10-fold cross-validation.⁶⁸ To incorporate differences in citations for papers published in different time periods and disciplines, we fit the model

⁶⁶Hill and Stein (2021) use a similar approach based on information from the Protein Data Bank to train a machine-learning model to predict citations of academic research.

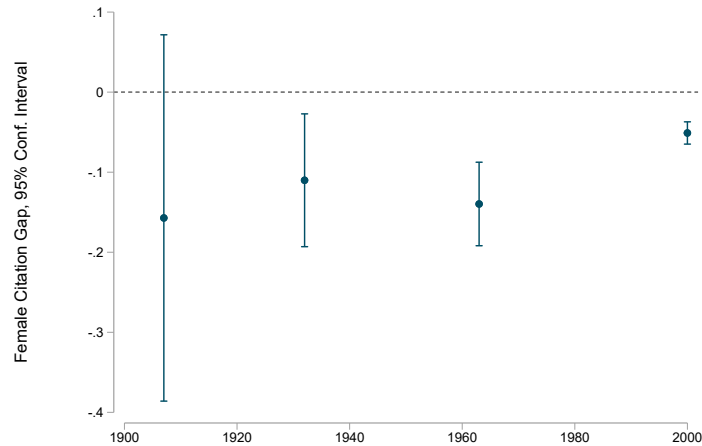
⁶⁷The last indicator equals one for all titles with 30 words or more.

⁶⁸We consider values of λ in a range of 1 to 10.

separately for each of our cohorts and disciplines. The model can thus account for the changing importance of topics over time and across disciplines.

D.2. Further Results: Citation Gaps

Figure D.1: Gender Gaps in Citations over Time



Notes: The Figure shows gender gaps in citations over time for the sample of prestigious universities (Sample 3). The gender gaps are estimated with equation (5). The regression controls for cohort-discipline-country fixed effects and the predicted citation controls.

E Further Results: Promotion Gaps

Table E.1: Promotions to Full Professor by Department Quality

Promotion to Full Professor	<i>Sample 2</i>			<i>Sample 3</i>		
	Men	Women	Total	Men	Women	Total
No Promotion	71.63%	87.83%	72.76%	69.85%	85.33%	70.64%
Promotion in Worse Department	4.71%	1.20%	4.47%	5.38%	1.72%	5.19%
Promotion in Same Department	20.20%	10.20%	19.50%	20.59%	12.17%	20.16%
Promotion in Better Department	3.46%	0.77%	3.27%	4.18%	0.78%	4.01%
Observations	15,482	1,167	16,649	12,013	641	12,654

Notes: The Table shows the probability of promotion to full professor by department quality and gender. Sample 2 includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Sample 3 includes scientists in prestigious universities until 2000. The quality of departments is determined from the ranking of the average (over scientists and across cohorts) standardized citations of departments as observed in Sample 2. The probabilities of promotion to full professor in a worse, same, or better department are computed as averages of indicators that equal 1 if a scientist who entered the dataset in cohort $t - 1$ at a lower rank than full professor was promoted to full professor by cohort t in a department of worse, same, or better quality than the quality of the scientist's department in cohort $t - 1$.