DISCUSSION PAPER SERIES

DP17422

Gender Gaps in Academia: Global Evidence over the Twentieth Century

Alessandro Iaria, Carlo Schwarz and Fabian Waldinger

ECONOMIC HISTORY

LABOUR ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

Gender Gaps in Academia: Global Evidence over the Twentieth Century

Alessandro Iaria, Carlo Schwarz and Fabian Waldinger

Discussion Paper DP17422
Published 01 July 2022
Submitted 30 June 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Gender Gaps in Academia: Global Evidence over the Twentieth Century

Abstract

Using the largest database of academics ever assembled, we analyze gender gaps in hiring, publications, citations, and promotions over an unprecedented time span and geographic coverage. First, we document an increasing share of female academics over the 20th century, in particular since 1970, with substantial heterogeneity across countries and disciplines. Second, we estimate gender gaps in publications of about -0.3sd. We uncover a U-shaped relationship of publication gaps as a function of the share of women in academia. A Roy model rationalizes this relationship and indicates persistent positive selection of women into academia throughout the 20th century. Third, we estimate negative gender gaps in citations. To control for gender differences in research topics, we develop a novel machine learning approach that predicts the citations of each paper, as if it had been written by men. Fourth, we estimate negative gender gaps in promotions, which persist even controlling for publications.

JEL Classification: N/A

Keywords: gender gaps, Universities, selection, Roy model, Predicted Citations

Alessandro Iaria - alessandro.iaria@bristol.ac.uk
University of Bristol and CEPR

Carlo Schwarz - carlo.schwarz@unibocconi.it
Bocconi University and CEPR

Fabian Waldinger - fabian.waldinger@econ.lmu.de
University of Munich and CEPR

Gender Gaps in Academia: Global Evidence over the Twentieth Century

Alessandro Iaria (University of Bristol)

Carlo Schwarz (Bocconi University)

Fabian Waldinger (University of Munich)*

June 30, 2022

Abstract

Using the largest database of academics ever assembled, we analyze gender gaps in hiring, publications, citations, and promotions over an unprecedented time span and geographic coverage. First, we document an increasing share of female academics over the 20th century, in particular since 1970, with substantial heterogeneity across countries and disciplines. Second, we estimate gender gaps in publications of about -0.3sd. We uncover a U-shaped relationship of publication gaps as a function of the share of women in academia. A Roy model rationalizes this relationship and indicates persistent positive selection of women into academia throughout the 20th century. Third, we estimate negative gender gaps in citations. To control for gender differences in research topics, we develop a novel machine learning approach that predicts the citations of each paper, as if it had been written by men. Fourth, we estimate negative gender gaps in promotions, which persist even controlling for publications.

*We thank Ran Abramitzky, Oriana Bandiera, David Card, Amy Finkelstein, Ben Jones, Muriel Niederle, Claudia Olivetti, Nina Roussille, and seminar participants at Berkeley, Bocconi, Boston University, Boulder, Bristol, Florida, Frankfurt, Munich, NYU, Stanford, Turin, UCLA, Warwick, York, CEPR 2022, and Barcelona Summer Forum 2022 for useful 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, Peter Heinrich, Robert Johannes, Simon Kallmaier, Christoph Kuehn, Paulina Maier, Axel-Konstantin Marko, Lukas Mergele, Luisa Neumeier, Tobias Nowacki, Adrian Nutzinger, Franz Oertel, Julia Ostendorf, Nicolas Porschke, Joseph Prestwich, Akos Reitz, Tabea Rudolf, Riyan Sawalha, Constanze Schwarz, Nils Süßenbach, Paul von Tein, 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.

Until the beginning of the 20th century, high-skilled professions were almost exclusively occupied by men. Even today, women remain under-represented in most high-skilled professions, especially in senior positions. For example, in 2017, women held only 5.8% of CEO positions in Fortune 500 firms (Bertrand et al., 2019) and were responsible for only 9.5% of patented inventions in OECD countries (OECD, 2021). In academia, women remain under-represented in most disciplines and countries. The persistent under-representation of women is of increasing concern for researchers, policymakers, and the general public (e.g., Beidas et al., 2022, Romanowicz, 2019). In recent years, a growing literature has analyzed gender gaps in academia investigating either publications or citations, and focusing on specific disciplines, countries, and time periods.

Despite these endeavors, much remains to be understood about gender gaps. In particular, little is known about the long-run evolution of gender gaps in academia across countries and disciplines. Furthermore, it remains unclear in which domains (hiring, publishing, citation behavior, promotions) gender gaps are more pronounced and if gender gaps interact across these domains.

This paper attempts to make progress in our understanding of gender gaps in academia by tracing their evolution across multiple domains over the 20th century and at the global level. The analysis leverages the largest database of university academics ever assembled. We hand-collect the data from historical records and modern university websites. The database includes about half a million observations covering academics in 7,484 universities in all disciplines in more than 130 countries for six cross-sections (cohorts) spanning the years 1900, 1914, 1925, 1938, 1956, and 1969.[1] We augment these data with information on academics in three academic disciplines for the year 2000, covering prestigious universities for which we have information throughout the 20th century.

The unique quality of our data derives from a large number of manual enhancements that enrich the faculty rosters. We code the gender of academics in a multi-step procedure. Additionally, we follow academic careers by developing 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] Further, we manually recode more than 100,000 specializations into 36 disciplines. For example, the specializations "advanced reactor theory and quantum theory" or "physique des particules élémentaires" are assigned to "physics." We also manually harmonize academic

[1] For comparison, the U.S. News ranking of worldwide universities is based on a pool of 1,748 universities (see USNEWS/Methodology, accessed August 6, 2021). The Shanghai Ranking of World Universities includes 2,417 universities (see https://www.shanghairanking.com/institution, accessed August 6, 2021).

[2] Margarete Bieber, an archaeologist and art historian, was the second woman in Germany to become full professor. Because of her Jewish background, she was dismissed by the Nazi government and emigrated to the United States (Becker et al., 2021).

ranks across countries, e.g., professor, associate professor, assistant professor to enable us to study academic promotions. Finally, we complement the faculty rosters with publication and citation data from *Clarivate Web of Science* and *Microsoft Academic Graph* to study gender gaps in publications and citations.[3]

Because the data are based on complete faculty rosters rather than publication databases, we can study the entire population of academics who are, in principle, able to publish or be promoted to full professor. This helps us overcome important selection biases and draw a more accurate portrait of the role of women in academia, which is not restricted to the sample of publishing academics.[4] Additionally, our data allow us to study gender gaps in four domains: hiring, publications, citations, and promotions. Each part of the paper studies gender gaps in one of these four domains and establishes new facts about their global evolution over the 20th century. We also investigate how gaps interact across domains.

In the first part of the paper, we document gender gaps in hiring. For many disciplines, countries, and universities, the data cover the first women to enter academia and the first women ever to be promoted to full professor, e.g., Katherine Ellis Coman, the first female full professor in economics in the United States (Vaughn, 2004). Across all universities and disciplines, our newly collected data show that in 1900 only 226 women had been hired, a share of about 1 percent. In the following decades, the share of women in academia increased slowly: 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, chemistry, biochemistry), female shares across all universities increased from about 1 percent to about 6.6 percent between 1900 and 1969. In prestigious universities, female shares in the sciences were about 25-50 percent lower (than in all universities) until 1969. Between 1969 and 2000, female shares in the sciences in prestigious universities increased substantially: from about 3.3 percent to 17.6. Despite the increase over the last decades of the 20th century, women are still widely underrepresented in prestigious universities. Our data also indicate that throughout the 20th century, female participation in academia was remarkably lower than in the general labor market for almost all countries.

We further investigate how gender gaps in hiring varied across academic ranks and document particularly large gaps for full professors. By 1900, all universities across the globe combined had only hired 113 women as full professors, a share of around 1 percent. In the following decades, the share of female full professors increased but always remained below the share among all academics. The slower increase in the share of women among full profes-

[3]As always with data on a global scale, comparability across countries and over time raises challenges. Section 1 discusses several robustness tests for each of our empirical findings.

[4]A recent paper studying gender inequality in scientific careers based on publication data is Huang et al., 2020.

sors could either reflect compositional changes over time or worse career prospects for female academics. We investigate this question at the end of the paper.

The global coverage of the data enables us to uncover substantial heterogeneity in the share of female academics across countries. Before WWI, universities in the United States hired more female academics than any other country, both in absolute and in relative terms. Overall, among all universities and disciplines, the dominant role of the United States persisted until 1969. Among the prestigious universities, the early lead of the United States in the sciences did not last, and by 2000 many other countries had hired more women in their prestigious universities. The United Kingdom also started the 20th century with a relatively high female share, but fell behind by 2000. In contrast, Scandinavian countries, and to a lesser extent Germany, had very low female shares until 1969 but increased their shares substantially in the three decades until 2000. Italy had low shares before WWI, was ranked in the middle until 1969, but increased its share substantially in the three decades until 2000. Japan is a clear outlier: female shares were at similar levels during the first decades of the 20th century but, unlike the other advanced countries, did not show a marked increase until 2000. The cross-country evolution of female shares in academia was remarkably different from the evolution of female shares in the general workforce (e.g., Olivetti and Petrongolo, 2016). This suggests that women's careers in academia were affected by different factors than in lower-skilled professions.

Next, we uncover substantial heterogeneity in female shares across disciplines. Averaged over the period 1900 to 1969, no discipline had a female share greater than 35 percent. Most disciplines had female shares below 10 percent. Disciplines with particularly high female shares were pedagogy, communication studies, languages, sports sciences, and social sciences. Most of these disciplines were not research oriented but focused on teaching to undergraduates. Disciplines with particularly low female shares (below 2 percent averaged until 1969) were law, veterinary medicine, architecture, theology, and engineering.

The second part of the paper investigates gender gaps in academic output, as measured by publications. One of the unique advantages of studying academics is the availability of output measures (i.e., publications) that are comparable across time and space. Publications have been one of the key metrics to evaluate academics. Of course, publications do not measure the true ability of academics as they are influenced by preferences, discrimination, and other biases. Because our data are not limited to the sample of publishing academics, our analysis overcomes important selection concerns when comparing publications across gender. We measure publications over a $\pm$ 5 year interval around each cohort (e.g., 1995-2005 for scientists observed in 2000). In all universities, during the period 1900-1969, women published on average two to three fewer papers than men (or around 0.3 of a standard deviation) over the ± 5 year interval. The gender gap in publications is similar if we compare men and women in the same department and cohort, e.g., chemistry at Harvard in 1969. In prestigious universities, we estimate around 50 percent larger gender gaps in publications.

We also estimate the evolution of gender gaps in publications over the 20th century. The publication gap was small and insignificantly different from zero in 1900. In the following decades, the gender gap in publication increased dramatically and reached a maximum of 0.45 standard deviations by 1969. After that, the gap declined to about 0.2 standard deviations by 2000, indicating that gender gaps in publications persisted into the 21st century.

Our data also allow us to link the evolution of gender gaps in publications to the hiring patterns documented in the first part of the paper. The results suggest a U-shaped relationship between the gender gap in publications and the share of women in academia. Publication gaps were small in periods with very low shares of female scientists. This can be referred to as the "Marie-Curie" period: only exceptional women were hired, and, on average, they published as much as men despite potential biases in the publication market.[5] With increasing shares of women in the profession, gender gaps in publications turned negative. However, when the share of women increased beyond very low levels, the negative gender gaps in publications started to narrow.

We outline a model along the lines of Roy (1951) to interpret this relationship. 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 features lead a scientist's publication outcome to be a function of the share of female academics because of (a) *indirect* effects of selection and gender bias in the hiring market and (b) *direct* effects of gender bias in the publication market. Based on the model, we estimate a regression that rationalizes the U-shaped gender gap in publications. The model and our estimation results suggest that gender gaps are inherently connected: gaps in hiring may have indirect repercussions on the observed gender gaps in publications.[6] The estimation results indicate that women faced persistent direct gender bias in the publishing market, and that women have been positively selected into academia throughout the 20th century. This selection was particularly pronounced in countries and periods with low female shares in academia. Ashraf et al. (2022) document a similar positive selection of women into working for a large multinational firm in countries with low female labor force participation.

In the third part of the paper, we explore whether papers published by women received fewer citations. Our key contribution to the understanding of citation gaps is a novel machine learning approach to investigate whether potential citation gaps stem from gender differences

[5] See also Rossiter (1982), p. 130, who describes that in the early part of the 20th century female scientists "had to be not only better than the men [...] but, preferably, 'Madame Curies' [to deserve a place in science]."

[6] Mulligan and Rubinstein (2008) suggest a similar mechanism for general labor market trends.

in research topics. We train a regularized regression model that uses the words in paper titles to predict the expected number of citations for each paper. To prevent the model from internalizing biases against papers published by women, the training sample solely consists of papers published by men. Therefore, the model predicts the citations of each paper as if it had been written by men. The results indicate that, before WWI, papers by female authors received around 0.3 standard deviations fewer citations than papers by male authors. This gender gap holds even after controlling for the predicted citations of the paper, suggesting that citation gaps were not driven by women publishing on topics that generally received fewer citations. In the interwar period, papers by female authors received around 0.14 standard deviations fewer citations, a gap that declined to around 0.08 standard deviations in the 1950s and 1960s. By 2000, the gender gap in citations had declined to practically zero.

In the fourth part of the paper, we investigate how women progressed in their academic careers by studying gender gaps in promotions. We find that women were around 20-30 percentage points less likely to be promoted to full professor until the late 1960s. Strikingly, gender gaps in promotions remain similar if we control for the publication and citation record of the scientist. This unexplained gender gap in promotions was larger than the effect of a five standard deviations worse publication record. By the year 2000, the gender gap in promotions in the sciences had closed. Notably, gender gaps in promotions are very similar if we compare men and women who entered the data in the same department and cohort (e.g., chemistry in Berkeley in 1900).

In summary, we show the existence of significant gender gaps across four domains (hiring, publications, citations, and promotions) over the 20th century. However, our analysis indicates that gender gaps differ substantially across countries, disciplines, as well as domains and have changed over time. Our findings suggest that the substantial gender gaps in citations and promotions have declined to zero over the course of the 20th century. In contrast, gender gaps in hiring and publications remain large, even at the end of the 20th century. Our results also illustrate that gender gaps in different domains are interconnected. The documented barriers that excluded women from participating and contributing to scientific progress may have resulted in "lost Marie Curies," depriving the academic community of valuable ideas and potential breakthroughs.[7] In a world where ideas play an ever-increasing role, this will slow down scientific progress and ultimately economic growth (e.g., Romer, 1986; Romer, 1990; Jones, 1995a).[8]

[7]Bell et al. (2019) show that there are many "lost Einsteins" because some children have not been exposed to patenting by either parents, co-workers of parents, or neighbors.

[8]Previous research has shown that discrimination against women and Blacks in the wider economy lowered aggregate productivity in the United States (Hsieh et al., 2019).

The paper contributes to a growing literature on gender gaps in science and innovation. In economics, female-authored papers receive more citations than male-authored papers, suggesting that women need to overcome higher hurdles to publish (Card et al., 2020); women were less likely to be nominated as Fellows for the Econometric Society until the late 1970s but more likely to be nominated since the mid-2000s (Card et al., 2022); women receive less credit for group work (Sarsons, 2017; Sarsons et al., 2021); references to female-authored papers in economics are more likely to be omitted (Koffi, 2021); and female-authored papers have higher readability scores (Hengel, 2020). Covering subjects beyond economics, the literature has also documented that female research team members are less likely to be credited with authorship (Ross et al., 2022); that women had lower productivity while having young children during the first half of the 20th century (Moser and Kim, 2022); that women are more likely to volunteer and to be asked to volunteer for tasks with low promotability (Babcock et al., 2017); that a higher share of women in evaluation committees lowers promotion prospects of women (Bagues, Sylos-Labini, and Zinovyeva, 2017); and that physicians become disproportionately pessimistic about female surgeons' ability after a patient's death (Sarsons, 2019). We contribute to this work by providing a comprehensive analysis of gender gaps in academia, studying four important career outcomes (hiring, publications, citations, and promotions) over a large number of countries and throughout the 20th century.

Because we are able to study gender gaps across various domains, we can show that gender gaps in one domain (hiring, resulting in differential selection of men and women into academia) have repercussions on observed gaps in another domain (publications). This relates to recent findings showing that positive selection of women affects observed gender gaps in wages in a large multinational firm (Ashraf et al., 2022) and in the broader U.S. economy (e.g., Mulligan and Rubinstein (2008) and Hsieh et al., 2019).

Our work also contributes to the literature that analyzes gender gaps in certain high-skilled professions, e.g., MBA graduates (Bertrand, Goldin, and Katz, 2010), executives (e.g., Bertrand and Hallock, 2001; Gayle, Golan, and Miller, 2012; Albanesi, Olivetti, and Prados, 2015), lawyers (Azmat and Ferrer, 2017), pharmacists (Goldin and Katz, 2016), and engineers (Roussille, 2022); all in the United States, or more broadly college graduates in the United States (Black et al., 2008) or Sweden (Albrecht, Björklund, and Vroman, 2003). Our new database enables us to trace the evolution of gender gaps for one high-skilled profession at the global level and over the entirety of the 20th century. In contrast, most existing papers have analyzed one country and relatively limited time periods due to a lack of comparable data.[9] A nuanced description of gender gaps sheds light on the many failures and the few success

[9] In addition, an extensive literature has studied gender gaps in hiring and wages in the general workforce (see Altonji and Blank, 1999, Bertrand, 2011, Blau and Kahn, 2017, and Bertrand and Duflo, 2017 for surveys). Most of this earlier work has studied individual countries and limited time periods. A notable

stories of promoting female careers in academia. This may ultimately allow to improve the design of anti-discriminatory policies and help overcome barriers that deprive academia, and society, of some of the best minds and ideas.

1 A New Database of University Academics

At the heart of this paper is the largest database of university academics ever assembled. We hand-collect faculty rosters from the historical publication *Minerva Jahrbuch der Gelehrten Welt* and modern university websites. We combine these data with detailed publication and citation records from the *Clarivate Web of Science* and *Microsoft Academic Graph*. Throughout the paper, we present results for three main samples:

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

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*. In a time before the Internet, *Minerva* was the most important worldwide directory of academics. The publishers of *Minerva* contacted ministries of education, university administrators, and academics to ensure almost comprehensive coverage.[10]

Minerva was published in volumes containing cross-sections of academics. We digitize six volumes that cover the years 1900, 1914, 1925, 1938, 1952/56, and 1966/1969 (see Figure 1 for a sample page).[11] For the remainder of the paper, we refer to these years as cohorts. For

exception regarding the time period is Goldin's seminal research on gender gaps in wages and employment in the United States from the late 19th century until today (e.g., Goldin, 1989, 1990).

[10]For example, an article in *Nature* compared the French publication *Index Generalis: Annuaire Général des Universités* to *Minerva* and noted that "[i]n scope, as indicated by the sub-title, this annual is akin to the well-known 'Minerva Jahrbuch der gelehrten Welt'. It is, however, very much less exhaustive" (Nat 1930). To the best of our knowledge, there are no comparable data covering academics on a worldwide scale over many decades. To provide evidence of its coverage, we benchmark the *Minerva* data to smaller datasets that cover some universities and time periods. The benchmarking exercises suggest that *Minerva* indeed covered a large fraction of the world's academics (see Appendix A.5 for details).

[11]As the number of academics increased dramatically over time, *Minerva* published the last two cohorts in two installments. We refer to these cohorts using the later year, e.g., 1956 for the 1952/56 publication.

the digitization, we scan all pages of the relevant volumes and process them using optical character recognition software (OCR). In the next step, we extract all relevant information from the largely unstructured OCR output and hand-check each entry to remove spelling errors in names.

Minerva lists academics from all disciplines and thousands of universities in more than 100 countries. The data include 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 Collegio* in Rome. Virtually all Ph.D. granting institutions are included in the data. For example, the data contain academics in 1,575 universities in the United States, 305 universities in the United Kingdom, 318 in Germany, and 348 in France.[12]

Minerva lists the name of the university and the city and country, followed by faculty rosters.[13] For most universities, the data list assistant, associate, and full professors, but also honorary professors, and in some cases research positions and teaching positions. The faculty rosters usually report the name of each academic as well as a finely grained specialization. Overall, the faculty rosters from *Minerva* contain around half a million person-cohort observations (Table 1, sample 1) in 7,483 universities in more than 130 countries.

Modern Faculty Rosters For The Year 2000

For the year 2000, we digitize faculty rosters from archived university websites (available from the Internet Archive *Wayback Machine*). We focus on three academic disciplines: mathematics, chemistry, and biochemistry and collect faculty rosters for 242 prestigious universities in 34 countries. These universities reported faculty rosters in all six historical *Minerva* cohorts or are ranked in the top 100 places of the Shanghai ranking in 2020.[14] Between 1900 and 1969, these 242 institutions employed around 48 percent of all academics in mathematics, chemistry, and biochemistry. The academics at these prestigious universities also published around 69 percent of all papers.

[12]Compared to existing research in economics, our data contain more academics in a larger number of universities. For example, the notable data collection effort by De la Croix et al. (2020) contains 33,726 academics in 207 universities covering the period 1000 to 1800.

[13]For some lesser-known universities, especially in India, the source only reports the number of professors without listing their names. Furthermore, for some universities the source lists the names of professors but only reports the number of teaching positions (e.g., “10 lecturers”) without listing names. Across all cohorts, the source list 498,527 faculty members with names (this sample forms the basis for our analysis, Table 1) and 108,398 additional faculty members (e.g., the 10 lecturers) without names.

[14]The sample contains 69 universities in the United States, 30 in Germany, 22 in the United Kingdom, 21 in Italy, 18 in France, 9 in Switzerland, 7 each in Australia, Austria, and Canada, 5 in Belgium, 4 each in Denmark, Ireland, the Netherlands, and Sweden, 3 in Japan, 2 each in Argentina, Finland, Hungary, India, New Zealand, Portugal, Russia, Spain, and Yugoslavia/Serbia/Croatia, and 1 each in Bulgaria, the Czech Republic, Greece, Israel, Norway, Poland, Romania, Russia, Singapore, and Uruguay.

Enhancements of Faculty Roster Data

We make a large number of manual enhancements to the faculty roster data (see Appendix A.1 for details). First, we manually recode thousands of different university ranks (e.g., "professor," "chargé de cours," or "incaricato") into ten comparable categories (e.g., assistant professor, full professor, emerita/us, or teaching position; see Appendix A.1.1). Second, we manually recode over 100,000 different specializations (e.g., "quantum theory" or "physique des particules élémentaires") into 36 disciplines (e.g., chemistry, economics, law, theology, or history; see Appendix A.1.2). Third, if academics hold multiple positions within the same city or university (e.g., a double appointment in two departments), we combine the information into a single observation (see Appendix A.1.3). Fourth, we link academics across the seven cohorts using a cascading procedure (see Appendix A.1.4). Fifth, for academics who are only listed with their surname and initials (instead of the complete first name), we conduct a manual web search to find their complete first name (see Appendix A.1.5).[15] Sixth, we construct consistent university identifiers by linking universities across cohorts and tracking mergers and splits over the 20th century.

Identifying the Gender of Academics

We develop a five-step procedure to identify the gender of academics on a global scale. First, whenever available, we use the information on gender from the faculty rosters in *Minerva* (e.g., names preceded by Mlle., Lady, Lord, Cardinal) or from the department websites (mostly pictures but also personal pronouns in research descriptions). In all further steps, we rely on first names to identify the gender of academics.

In the second step, we process more than 100,000 'first name'-country combinations with *gender-api.com*,[16] which assigns a gender probability to 'first name'-country combinations.

In the third step, two research assistants (one male and one female) independently classify 'first name'-country combinations that *gender-api.com* classified as less than 100% male. The research assistants are instructed to only classify cases for which they can assign gender with certainty. If the two research 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 combination using a *Google* image search. A research assistant then classifies the 'first name'-country combination as male or female depending on whether the image search returns more male or female individuals. E.g.,

[15] All results remain unchanged without this step.

[16] *Gender-api.com* is a commercial solution with the key advantage of differentiating the gender of first names at the country level (e.g., Andrea is a male name in Italy but a female name in many other countries). *Gender-api.com* is currently the best performing name-to-gender inference service (Santamaría and Mihaljević, 2018).

gender-api and the research assistants could not identify the gender of "Hadmar" in Austria. We therefore search for "Hadmar Austria" in *Google* and analyze the images that *Google* returns. In this example, the images that depict individuals show only men (see Appendix Figure A.1), we therefore code the 'first name'-country combination as male.

In the fifth step, we hand-check individual academics who appear to be misclassified with an extensive *Google* search (see Appendix A.2.2).[17] Such misclassifications mostly occur because the predominant gender of some first names changed over time. E.g., the French name "Camille" can be both male and female. In the early cohorts, most academics with the first name "Camille" are male, while in later cohorts some are female. While the manual steps significantly increase data quality, none of the results change without steps 3 to 5. For the main results, we focus on the sample of academics for whom we can assign gender (see Table 1).

Table 1: Summary Statistics

	All Academics	All Gender Coded	Female	Male
Sample 1: All Universities; all disciplines; 1900-1969				
Number of academic - cohort observations	498,527	410,542	35,447	375,095
Number of universities	7,483	5,508	2,404	5,136
Number of departments	36,977	30,561	8,754	28,993
Female %		8.63	100.00	0.00
Sample 2: All Universities; math, chemistry, biochemistry; 1900-1969				
Number of academic - cohort observations		35,178	1,806	33,372
Number of universities		1,893	644	1,821
Number of departments		3,548	895	3,394
Female %		5.13	100.00	0.00
Publications		4.65	1.39	4.83
Sample 3: Prestigious Universities; math, chemistry, biochemistry; 1900-2000				
Number of academic - cohort observations		40,226	4,510	35,716
Number of universities		242	238	242
Number of departments		686	533	684
Female %		11.21	100.00	0.00
Publications		8.13	5.08	8.51

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, chemistry, and biochemistry from 1900 until 1969. Sample 3 includes academics in prestigious universities in mathematics, chemistry, and biochemistry from 1900 until 2000. The data were collected by the authors from various volumes of *Minerva*, university websites, *Clarivate Web of Science*, and *Microsoft Academic Graph* see section 1 for details.

[17]The *Google* search in this step does not only use a 'first name'-country combination but an actual academic with surname, first name, and discipline.

Figure 1: Examples of Academics in Database

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

Examples of Academics in the Database

Figure 1 shows three exemplary academics for each cohort. Choosing examples among more than half a million academic-cohort observations leads to somewhat arbitrary decisions. The selection showcases some of the data's country, discipline, cohort, and gender dimensions. However, it will not do justice to the tens of thousands of academics that 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 the economist John Maynard Keynes (University of Cambridge), the physicist and Nobel Laureate 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 achievement, 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).[18]

In 1925 the data contain the physics Nobel Laureate Niels Bohr (University of Copenhagen), the founder of psychoanalysis Siegmund Freud (University of Vienna) and Lise Meitner (University of Berlin). Lise Meitner was born in Vienna and became the second woman to earn a physics PhD at the University of Vienna. During her Post-Doc years at the University of Berlin, she did not receive a salary and had to run her experiments in a converted carpenter's shop in the basement because — as a woman — she was not allowed to enter the main building of the 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 Swedish Academy awarded the Nobel Prize to Hahn but overlooked Meitner's contribution. Contemporaries described the decision to omit Meitner 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. Joliot-Curie was Pierre and Marie Curie's daughter and 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

[18]Strikingly, Marie Curie is listed in *Minerva* as Mme P[ierre] Curie, at a time when she had won two Nobel prizes. Most of Marie Curie's papers were published under the name Mme P. Curie. During her entire career, Marie Curie faced obstacles because of her gender. In 1911, she was nominated to the French Academy of Sciences. Her nomination met strong resistance: "Women cannot be part of the Institute of France" argued the physicist Émile Amagat. Despite the efforts of some of France's greatest scientists, she lost the membership election by one vote to her male competitor (see Curie, 1938, pp. 277 for a detailed account).

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). Gerty 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 both of them an 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 philosopher John Rawls (Harvard), the biologist and discoverer of the double helix structure of the 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, she was 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 chemist Frances H. Arnold (Caltech) who pioneered the use of directed evolution to create enzymes and was awarded the Nobel Prize in Chemistry in 2018, the biochemist Elisabeth Blackburn (UC San Francisco), who co-discovered telomerase and was awarded the Nobel Prize in physiology/medicine in 2009, and the mathematician Terence Tao (UCLA), who was awarded the Fields Medal in 2006.

1.2 Publication and Citation Data

To study gender gaps in publications and citations, we augment the faculty roster data with publication and citation data from *Clarivate Web of Science.* For any result based on publications and citations, we focus on three academic disciplines: chemistry, biochemistry, and mathematics. There are three reasons for this. First, these disciplines have particularly good coverage in the *Web of Science.* Second, they had already established a culture of publishing in scientific journals by 1900, and the publishing process was similar to today's. Third, the publishing process was highly international (Iaria, Schwarz, and Waldinger, 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, disciplines, and over time. This does not affect our estimate because we control for cohort $\times$ country $\times$ discipline fixed effects in all regressions.

We match academics with their publications using a cascading 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.[19] To harmonize affiliations across the faculty rosters and the *Web of Science*, we rely on *Google Maps API* (see Appendix A.3.2). This allows us to extract cities and countries for each of the hundreds of thousands of relatively unstructured affiliations. E.g., we extract the city "Cambridge" and the country "UK" from the affiliation "Cavendish Lab., Cambridge University, UK."

The matching is always based on the primary discipline of an academic (e.g., chemistry) to reduce the number of false positives. As the *Web of Science* only assigns disciplines (e.g., mathematics, chemistry, or general science) at the journal level, we develop a machine learning classifier to assign disciplines to individual papers (see Appendix A.3.3). Assigning the correct discipline is especially important for papers published in multi-disciplinary journals, such as *Nature* or *Science.* These papers could otherwise not be merged on the basis of a discipline. The classifier is based on a L2-regularized multinomial logit model. The model predicts a discipline for each paper, based on the unigrams, bigrams, and trigrams from the titles of the 59% papers which were published in journals that are assigned to only one discipline (e.g., *Acta Mathematica* which is uniquely assigned to mathematics).

We consider publications in a $\pm$ five-year window around the year of the corresponding cohort. E.g., for scientists from the 2000 cohort, we match papers published between 1995 and 2005.[20] In the rare cases that two or more scientists have identical names and work in the same discipline, we assign the paper proportionally to each scientist.[21]

[19]For 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, an alternative database *Microsoft Academic Graph* (*MAG*) contains the relevant affiliation. We therefore enrich the affiliation information with data from *MAG* (see Appendix A.3.2).

[20]We use a $\pm$ five-year window because scientists do not necessarily publish every year. As we use the surname and first name/initials to match publications and academics, a possible concern is that women change their surname after marriage. This would prevent us from merging all relevant papers to women within the $\pm$ five-year window. Reassuringly, the estimated gender gaps remain unchanged if we merge publications using a $\pm$ three-year window. Thus, the name change would have had to occur in the short window between reporting their name in *Minerva* or on their website and the year of the publication. This supports the view that name changes do not substantially bias our results. In addition to the short time differences in observing names in faculty rosters and publications, two additional factors may explain the absence of bias. First, the faculty rosters predominately list academics who are assistant, associate, or full professors and, hence, individuals who most likely were already married if they ever married. Second, marriage rates for female academics in the early part of our data are relatively low, e.g., 18 percent in 1921 and 26 percent in 1938 for scientists in the United States (Rossiter, 1982, p. 140).

[21]Results are robust in a sample of scientists with unique lastname-first initial-discipline combinations in each cohort: i.e., a sample of academics listed in the faculty rosters as unique in terms of lastname, first initial, and discipline in any university of the world (Table C.1).

2 Gender Gaps in Hiring

Hiring of Women Over Time

In the first part of the analysis, we investigate the evolution of gender gaps in hiring. We show results for the following samples: sample 1: all universities, all disciplines, 1900-1969; sample 2: all universities, sciences (mathematics, chemistry, biochemistry) 1900-1969*;* and sample 3: prestigious universities, sciences (mathematics, chemistry, biochemistry), 1900-2000.

Across all universities and disciplines, our newly collected data show that in 1900 only 226 women had been hired, a share of about 1 percent (Figure 2, panel a). In the following decades, the share of women in academia increased, in particular between 1925 and 1938, i.e., before WWII.[22] By 1969, a total of 17,211 women worked across all universities and disciplines, a share of about 11 percent — still nowhere close to equal representation.

We also explore changes in the number of women who held full professor positions. All over the world, full professor is the highest academic rank, which guarantees unique privileges and particularly high job security and salaries. In addition, full professor is the most comparable academic rank across different university systems. In 1900, only 113 women worked as full professors across all universities in our data, representing about 1 percent of the profession. In the following decades, the share of women among full professors increased, and by 1969 reached about 8 percent.

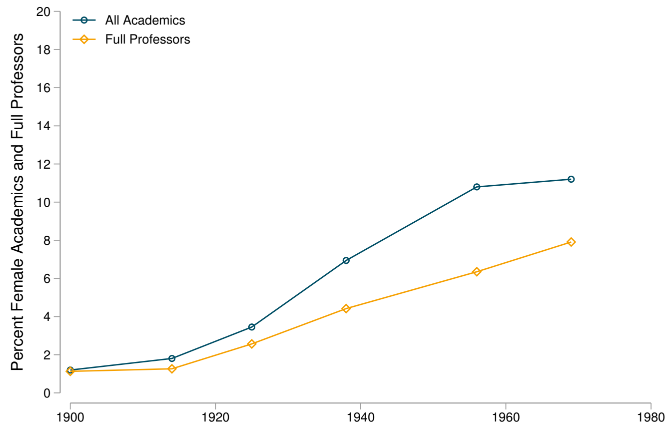
The slower increase in the share of women among full professors, compared to all academics, may indicate that women were less likely to be promoted but may also reflect compositional changes. For example, if a higher share of women was hired in later cohorts, it could take time for these women to rise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

In the sciences (mathematics, chemistry, biochemistry), female shares across all universities increased from about 1 percent to about 6.5 percent between 1900 and 1969 (Figure 2, panel b). In prestigious universities, female shares in the sciences were about 25-50 percent lower (than in all universities) until 1969. Between 1969 and 2000, female shares in the sciences in prestigious universities increased substantially: from about 3.3 percent to about 17.6 percent and from 2.2 percent to 7.7 percent among full professors (Figure 2, panel c). Despite this large increase, women are still heavily underrepresented in prestigious universities in 2000, especially among full professors.

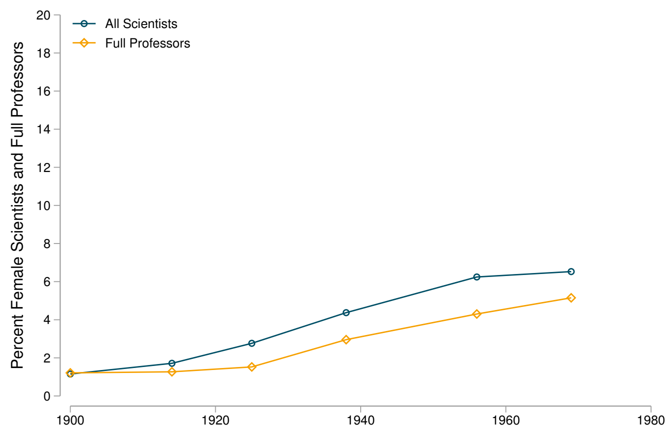
[22]The large increase of women in academia before WWII is quite distinct from general female labor force trends. In the United States, general labor force participation increased sharply during WWII (Acemoglu, Autor, and Lyle, 2004), but these trends did not persist, and many women returned to non-employment after WWII (Goldin, 1991).

Figure 2: Percent of Female Academics over Time

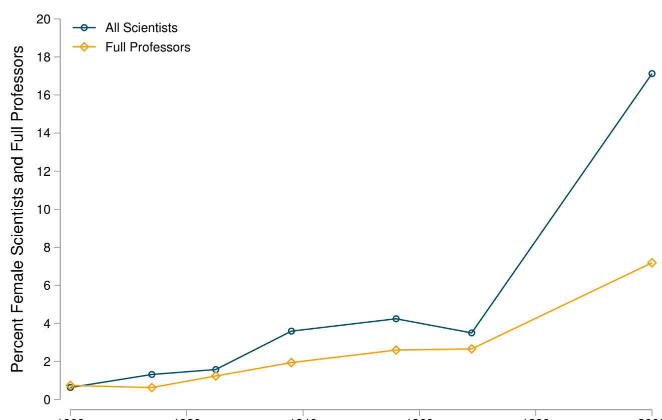
(a) *Sample 1:* All unis, all disciplines, 1900-1969



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



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



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

Hiring Gaps Across Countries

The aggregate statistics mask significant heterogeneity across countries. Before WWI, universities in the United States hired more women than any other country in the world, both in absolute and relative terms. In the sample of all universities and disciplines (sample 1), the dominant role of the United States persisted until 1969 (Figure 3, panel a).[23] In the sciences across all universities (sample 2), the United States continued to play a major role in hiring women until 1956 but was then overtaken by France (Figure 3, panel b). In prestigious universities (sample 3), the early lead of the United States in the sciences did not last, and by the year 2000 many other countries exhibit higher female shares (Figure 3, panel c).

The United Kingdom also started the 20th century with relatively high female shares but was overtaken by 2000. In contrast, Sweden, and to a lesser extent Germany, had very low female shares until 1969, but increased its share substantially in the three decades until 2000.[24] Italy had low shares before WWI, was ranked in the middle until 1969, but then increased its share substantially in the three decades until 2000.[25] Japan is a clear outlier: female shares were at similar levels during the first decades of the 20th century but, unlike the other advanced countries, did not show a marked increase until 2000.

Interestingly, the country-level patterns are distinct from trends in female employment shares among the general population (e.g., Olivetti and Petrongolo, 2016). This suggests that the careers of women in academia evolved differently from lower-skilled professions.

Hiring Gaps Across Selected Universities

Our detailed data also allow us to explore university-level variation in hiring gaps. It goes without saying that 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 figure). To select universities for this exercise, we rely on the well-known *Shanghai Ranking* of universities (see Ranking, 2020 for details). We choose the highest-ranked universities in each country and report the average female shares from 1900 to 2000.[26]

[23] The early U.S. lead is partly explained by women's colleges hiring more women. However, by 1938 also other U.S. universities were hiring more women than universities in other countries (Appendix Figure B.1a).

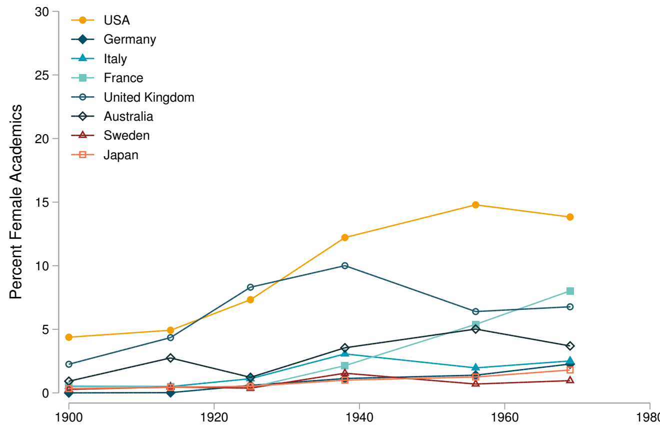
[24] Austria and Finland show a similar development, as shown in Appendix Figure B.1.

[25] Other Latin countries had a similar development. Female shares in 2000 were 51 percent in Argentina, 37 percent in Spain, and 28 percent in Portugal (Appendix Figure B.1).

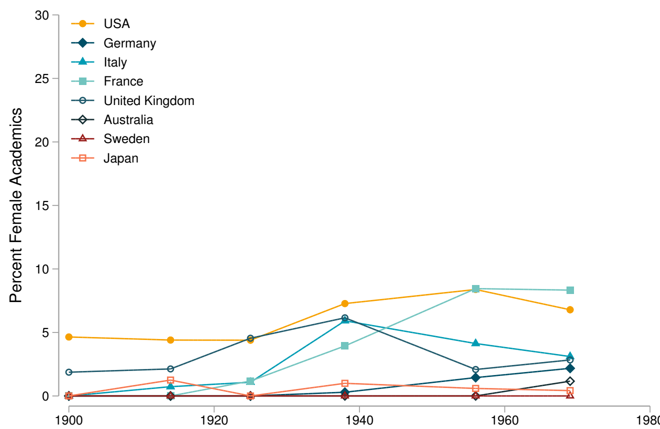
[26] The Shanghai Ranking ranks universities as of 2020. In many countries, e.g., the United States, the ranking of universities has remained stable since 1900. In other countries the ranking has changed substantially. To report the most important institutions over the 20th century, we deviate somewhat from the Shanghai ranking for Germany and France. For Germany, we include the *University of Berlin/Humboldt University*, the premier institution in Germany until WWII, instead of the *University of Bonn*. In France, several reorganizations of universities occurred during the 20th century. To capture some of the highest-ranked institutions for the entire 20th century, we include the *Université de Paris* and *Université de Grenoble*.

Figure 3: Percent of Female Academics by Country over Time

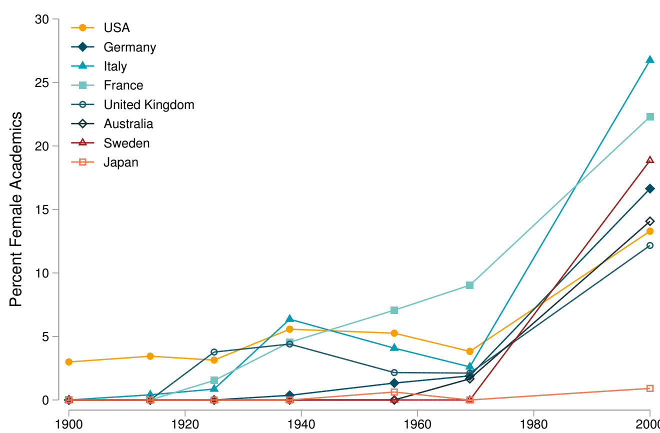
(a) *Sample 1:* All unis, all disciplines, 1900-1969



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



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

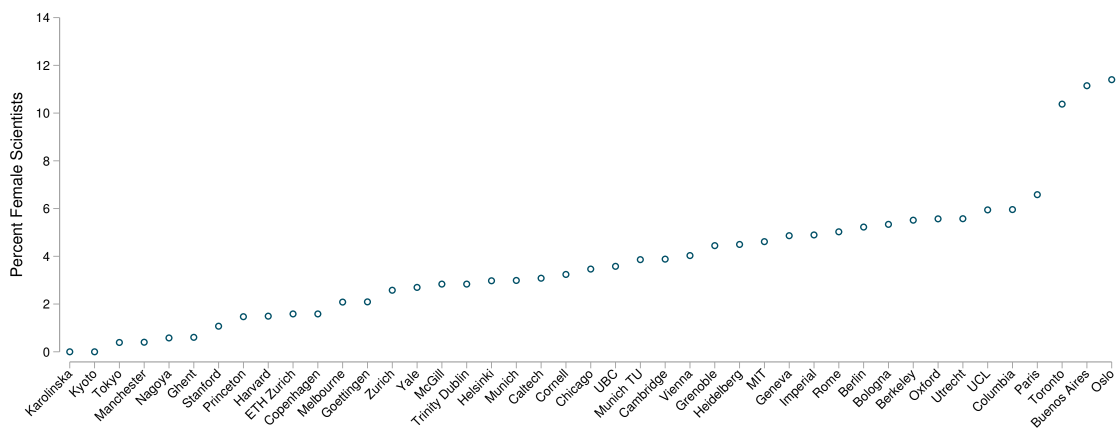


Notes: The Figure shows the percentage of female academics by country and over time. Panel a shows female shares in all universities and disciplines until 1969. Panel b shows female shares in all universities in the sciences (mathematics, chemistry, biochemistry) until 1969. Panel c shows female shares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

We report data on ten universities from the United States; five universities each from Germany and the United Kingdom, three universities each from Canada, Japan, and Switzerland; two universities from France and Italy; and one university each 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. E.g., over the 20th century, Berkeley hired on average 6 percent women in the sciences, while Stanford only hired 1 percent.

Figure 4: Percent of Female Scientists by University 1900-2000



Notes: The Figure shows the percentage of female scientists (mathmatics, chemistry, biochemistry) by university in the sample of prestigious universities (sample 3). Universities were selected as explained in the text. We calculate percentages of female academics at the cohort and university-level, e.g., Harvard in 2000, and then average the percentages over the seven cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort). The data were collected by the authors from various volumes of *Minerva* and department websites, see section 1 for details.

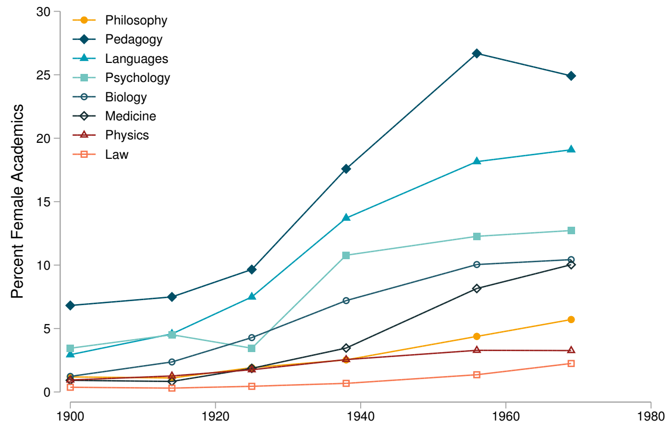
Hiring Gaps Across Disciplines

Our data also allow us to document differences in hiring gaps across disciplines (Figure 5). In the first decades of the 20th century, most disciplines had very low (below 5 percent) female shares. For most disciplines, female shares remained below 10 percent until 1969, with particularly low shares in law, physics, and philosophy. However, some disciplines had higher female shares, which increased to almost 25 percent by 1969 in pedagogy and about 20 percent in languages (Figure 5, panel a). Appendix Figure B.2 shows additional disciplines.

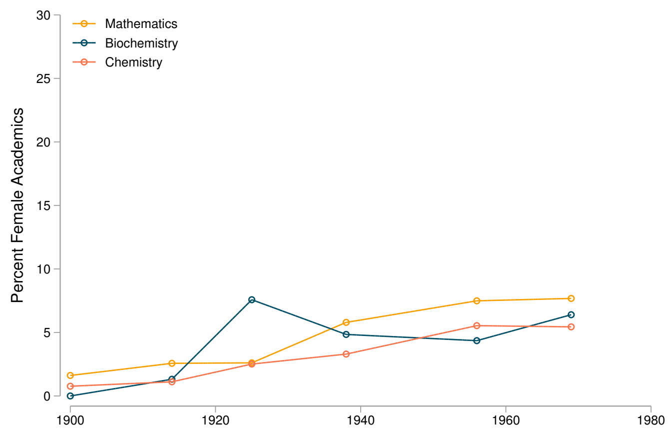
In the science sample (mathematics, chemistry, and biochemistry), the female share across all universities increased from close to 0 to between 5 and 7 percent by 1969. In the sample of prestigious universities, female shares were even lower until 1969. In the last three decades of the 20th century, they increased substantially and by 2000 reached around 14 percent in mathematics, 18 percent in chemistry, and 26 percent in biochemistry.

Figure 5: Percent of Female Academics by Discipline over Time

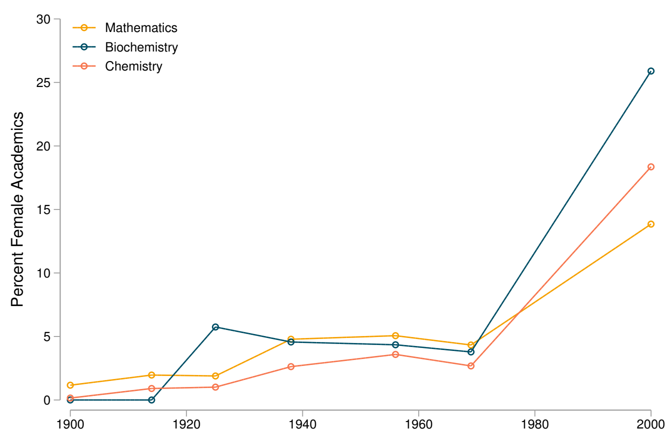
(a) *Sample 1:* All unis, all disciplines, 1900-1969



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



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



Notes: The Figure shows the percentage of female academics by discipline. Panel a shows female shares in eight disciplines in all universities until 1969. Panel b shows female shares in the sciences (mathematics, chemistry, and biochemistry) in all universities until 1969. Panel c shows female shares in the sciences (mathematics, chemistry, and biochemistry) in the sample of prestigious universities until 2000. The data were collected by the authors from various volumes of *Minerva* and department websites, see section 1 for details.

3 Gender Gaps in Publications

In this section, we explore gender gaps in publications. One of the unique advantages of studying academics is that we observe individual-level measures of output (publications) that are comparable across time and space. In contrast, comparable measures are usually not available in other industries. Publications are key metrics of performance that are used to evaluate career progression, allocate research funds, and rank academics. As previously discussed, we do not interpret publications as a measure of the true ability of academics. They reflect gender differences in output that could stem from differences in preferences, discrimination in the journal market or in the workplace (e.g., because women had worse access to high-quality labs), and other gender imbalances (e.g., differences in childcare at home).

To estimate gender gaps in publications, we focus on academics working in disciplines for which we have detailed publication and citation data (mathematics, chemistry, and biochemistry). We estimate "Mincer-type" regressions for sample 2: all universities 1900-1969 and sample 3: prestigious universities 1900-2000:

$$\begin{aligned}
\text{Pub}_{idt} &= \beta_1 + \beta_2 \text{Female}_{idt} \times 1\left[t(i) = 1900/14\right] + \beta_3 \text{Female}_{idt} \times 1\left[t(i) = 1925/38\right] \\
&+ \beta_4 \text{Female}_{idt} \times 1\left[t(i) = 1956/69\right] + \beta_5 \text{Female}_{idt} \times 1[t(i) = 2000] \qquad (1) \\
&+ \text{Experience}_{idt} \boldsymbol{\beta}_6 + \text{Fixed Effects} + \varepsilon_{idt}
\end{aligned}$$

where Pub_{idt} measures the number of papers that scientist i in cohort t and department d (e.g., mathematics at Harvard, which also determines i's country and discipline) published in the journals covered by the *Web of Science*.[27] As described above, we measure scientist i's papers in a $\pm$ five-year window around i's cohort $t(i)$. I.e., for scientists that we observe in 2000, we consider papers published between 1995 and 2005. The main explanatory variables are the interactions of the indicator variable Female_{idt} with indicators for four different time periods: pre-WW1 (1900 and 1914 cohorts), interwar (1925 and 1938), post-WW2 (1956 and 1969), and modern (2000). All regressions control for experience, which we measure as the number of times the scientist was observed in our data.[28] We also control for a large

[27]A small proportion of scientists (around 5%) have more than one affiliation in the same city and cohort, either in multiple departments of the same university or across universities. E.g., in 1914, the Russian-Italian chemist Maria Bakunin held appointments at the University of Naples and the Technical University of Naples. She was part of a group studying the eruption of Mount Vesuvius. To avoid double-counting, we estimate all our regressions using only one observation for each academic and cohort. We obtain almost identical results in regressions that keep all affiliations for each scientist (i.e., including a scientist twice if she was affiliated with two universities) or if we drop all academics with multiple affiliations.

[28]We include a set of variables that indicate the number of times the scientist has been observed before the relevant cohort. For example, a scientist who entered the data in 1956 and remained until 1969 contributes

number of additional fixed effects. In the baseline specification, we control for the three-way interaction of cohort, discipline, and country fixed effects (e.g., a separate fixed effect for mathematics in the Unites 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 sets of fixed effects. To account for the potential correlation of the residual ε_{idt} within country-discipline cells (e.g., chemistry in the United States), we cluster the standard errors at the country-discipline level.

The 1900 and 1914 cohorts of female scientists published on average 1.9 fewer papers than men in the full sample of all universities (sample 2). The 1925 and 1938 cohorts of female scientists published 2.4 fewer papers, and the 1956 and 1969 cohorts published 2.9 fewer papers (Table 2, sample 2, column 1, all significant at the 1 percent level). These are substantial gaps compared to the mean of publications, which is around 4.7 over the whole sample period. Even comparing women to men within the same department (sample 2, column 2), the publication gaps only shrink slightly. Controlling for departments is similar to controlling for industries or occupations in standard Mincer regressions. Note, however, that department choice is potentially endogenous. Finally, in column 3, we control for cohort $\times$ department fixed effects. We thus estimate publication gaps for scientists in the same department and cohort (e.g., Harvard chemists in 2000). Even within this restricted comparison group, we find a similar pattern of gender gaps in publications over time.

The coverage of journals in publication databases 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. E.g., many women entered the data in later periods and worked in the United States, i.e., a period and country with higher average publications. We therefore show alternative specifications that use normalized publications as the dependent variable. We normalize the number of publications to have mean 0 and standard deviation 1 within each country, cohort, and discipline (e.g., mathematics in the United States in 1969). Using this measure, we find a negative gender gap in publications of around 0.2 to 0.3 of a standard deviation (Table 2, sample 2, columns 4-6, all significant at the 1 percent level).

two observations. The experience indicator corresponding to observing the scientist for the first time is equal to 1 for the observation in 1956 but equal to 0 for the observation in 1969. The experience indicator corresponding to observing the scientist for the second time is equal to 0 for the observation in 1956 but equal to 1 for the observation in 1969. Experience indicators for observing the scientist a third, fourth, or fifth time are all zero in this particular case. Results are very similar in samples that restrict the data to the first observation for each scientist (Table 3), suggesting that gender gaps in publications are not driven by the fact that women are observed at different career stages than men.

Table 2: Gender Gaps in Individual-Level Publication

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:		Publications			Standardized Publications	
Sample 2: All Universities 1900-1969						
Female (1900/14)	-1.922***	-0.368	-1.644**	-0.322***	-0.142***	-0.234***
	(0.539)	(0.370)	(0.707)	(0.030)	(0.045)	(0.058)
Female (1925/38)	-2.433***	-1.868***	-2.431***	-0.287***	-0.249***	-0.302***
	(0.810)	(0.603)	(0.751)	(0.028)	(0.041)	(0.045)
Female (1956/69)	-2.874***	-2.022***	-1.884***	-0.281***	-0.193***	-0.182***
	(0.881)	(0.628)	(0.639)	(0.024)	(0.022)	(0.031)
Observations	35,178	35,178	35,178	35,178	35,178	35,178
R-squared	0.174	0.289	0.349	0.005	0.153	0.238
Sample 3: Prestigious Universities 1900-2000						
Female (1900/14)	-2.340***	-1.800***	-2.129**	-0.373***	-0.239***	-0.263***
	(0.845)	(0.673)	(0.903)	(0.048)	(0.075)	(0.057)
Female (1925/38)	-3.879***	-3.466***	-3.791***	-0.432***	-0.374***	-0.402***
	(1.236)	(1.097)	(1.050)	(0.040)	(0.038)	(0.038)
Female (1956/69)	-5.032***	-4.901***	-4.133***	-0.482***	-0.450***	-0.389***
	(1.320)	(1.243)	(1.233)	(0.050)	(0.040)	(0.050)
Female (2000)	-4.172***	-3.523***	-3.251***	-0.238***	-0.205***	-0.201***
	(1.194)	(0.989)	(0.820)	(0.043)	(0.038)	(0.034)
Observations	40,226	40,226	40,226	40,226	40,226	40,226
R-squared	0.213	0.246	0.280	0.020	0.064	0.132
Experience	Yes	Yes	Yes	Yes	Yes	Yes
Cohort × Discipline × Country FE	Yes	Yes		Yes	Yes	
Department FE		Yes			Yes	
Cohort × Department FE			Yes			Yes

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist-level. Sample 2 includes scientists (mathematicians, chemists, and biochemists) in all universities covering the period 1900-1969. Sample 3 includes scientists in prestigious universities covering the period 1900-2000. In columns 1-3, the dependent variable equals the number of publications in a ± five-year window around a cohort (i.e., 1995-2005 for a scientist listed in 2000). In columns 4-6, the dependent variable equals the number of publications, but standarized at the country-cohort-discipline level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman. The regressions also control for experience and various additional fixed effects, as indicated in the table. Standard errors are clustered at the discipline-country level. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

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 throughout the 20th century. As a result, compositional changes in the sample of universities cannot drive the findings.

In this sample, the publication gaps are larger (in absolute magnitude) than in the unrestricted sample of universities (Table 2, sample 3). As scientists in these universities published, on average, more papers (Table 1), the publication gaps are similar in percentage terms. In prestigious universities, the gender gaps in publications increased from around

0.37 s.d. for the 1900 and 1914 cohorts to around 0.48 s.d. for the 1956 and 1969 cohorts. For the 2000 cohort, we estimate a gap of around 0.24 s.d., indicating that publication gaps have declined in the latter part of the 20th century.

Table 3: Individual-Level Publication Gaps (Robustness)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	±Three-Year Window		Unique Matches		Full Professors		First Cohort	
Dependent Variable	Publi-cations	Standard. Publi-cations	Publi-cations	Standard. Publi-cations	Publi-cations	Standard. Publi-cations	Publi-cations	Standard. Publi-cations
Sample 2: All Universities 1900-1969								
Female (1900/14)	-1.327***	-0.308***	-1.946***	-0.320***	-1.037***	-0.252***	-1.814***	-0.299***
	(0.386)	(0.028)	(0.555)	(0.031)	(0.331)	(0.035)	(0.495)	(0.030)
Female (1925/38)	-1.468***	-0.242***	-2.398***	-0.282***	-2.836***	-0.371***	-2.358***	-0.277***
	(0.523)	(0.027)	(0.823)	(0.031)	(1.053)	(0.038)	(0.839)	(0.038)
Female (1956/69)	-1.972***	-0.276***	-2.849***	-0.276***	-3.354***	-0.304***	-2.604***	-0.273***
	(0.599)	(0.025)	(0.894)	(0.027)	(1.020)	(0.054)	(0.768)	(0.024)
Observations	34,321	34,321	31,492	31,492	18,456	18,456	26,267	26,267
R-squared	0.174	0.005	0.172	0.007	0.210	0.018	0.180	0.010
Sample 3: Prestigious Universities 1900-2000								
Female (1900/14)	-1.595***	-0.348***	-2.322***	-0.363***	-1.015***	-0.244***	-2.237***	-0.340***
	(0.603)	(0.040)	(0.858)	(0.048)	(0.230)	(0.049)	(0.810)	(0.047)
Female (1925/38)	-2.296***	-0.348***	-3.809***	-0.429***	-3.817***	-0.535***	-3.773***	-0.395***
	(0.818)	(0.053)	(1.242)	(0.044)	(1.410)	(0.063)	(1.299)	(0.057)
Female (1956/69)	-3.472***	-0.475***	-4.918***	-0.470***	-6.676***	-0.538***	-4.811***	-0.519***
	(0.898)	(0.052)	(1.318)	(0.059)	(1.669)	(0.069)	(1.108)	(0.063)
Female (2000)	-2.675***	-0.224***	-4.124***	-0.242***	-0.776	-0.139**	-4.201***	-0.239***
	(0.763)	(0.041)	(1.181)	(0.041)	(1.637)	(0.060)	(1.199)	(0.044)
Observations	39,816	39,816	36,514	36,514	17,588	17,588	33,782	33,782
R-squared	0.209	0.019	0.208	0.022	0.297	0.038	0.218	0.024
Experience	Yes	Yes	Yes	Yes	Yes	Yes		
Cohort × Discipline × Country FE	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 (mathematicians, chemists, and biochemists) in all universities covering the period 1900-1969. Sample 3 includes scientists in prestigious universities covering the period 1900-2000. In columns 1-2, a scientist's publications are counted in a ± three-year window around a cohort (i.e. 1911-1917 for a scientist listed in 1914). In columns 3-4, each scientist is defined as a unique lastname - first initial - discipline combination in every cohort and a scientist's publications are counted in a ± five-year window. In columns 5-6, only full professors are included in the analysis. In columns 7-8, only the first cohort in which a scientist is observed in the data is included in the analysis. In columns 1, 3, 5, and 7 the dependent variable equals the number of publications, while in columns 2, 4, 6, and 8 the standarized number of publications at the country-cohort-discipline level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman. The regressions control for cohort×discipline×country fixed effects. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

Robustness

The estimated gender gaps in publications are robust to alternative ways of linking papers to academics. First, we show that results are similar if we measure publications over a ± three-year window around i's cohort. I.e., for scientists in the 2000 cohort, we consider papers published between 1997 and 2003. Naturally, the point estimates are lower because the mean number of publications is lower in a ± three-year window. (Table 3, columns 1-2). We also show that publication gaps are very similar in a sample of scientists with unique lastname -

first initial - discipline combinations in every cohort (Table 3, columns 3-4). These results suggest that gender differences in publications do not stem from gender differences in the frequency of certain lastname - first initial combinations. Publication gaps are also similar in a sample of full professors, the academic rank that is most comparable across countries (Table 3, columns 5-6). Finally, results are similar in a sample that only includes each scientist in the first cohort for which the scientist is observed in the data (Table 3, columns 7-8). Appendix Table C.1 shows that these results remain robust with additional fixed effects.

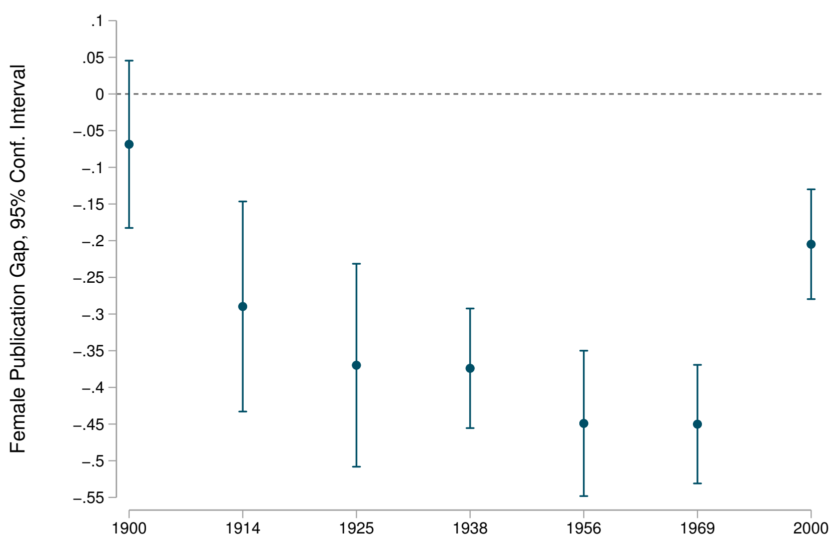
3.1 Publications Gaps Over Time

To study the evolution of gender gaps in publications over the 20th century at a finer level, we interact the female indicator with indicators for each cohort t:

$$\text{Pub}_{idt} = \beta_1 + \sum_{\tau=1900}^{2000} \beta_\tau \text{Female}_{idt} \times 1[t(i) = \tau] + \text{Experience}_{idt}\boldsymbol{\beta}_2 + \text{Fixed Effects} + \varepsilon_{idt}. \tag{2}$$

We then plot the seven coefficients $\hat{\beta}_\tau$ and corresponding 95 percent confidence intervals (Figure 6). The publication gap was small and insignificantly different from 0 in 1900. The gap increased until 1956 to a maximum of around 0.45 standard deviations and then started declining until 2000.

Figure 6: Gender Gaps in Publications over Time



Notes: The Figure shows gender gaps in publications over time for the sample of prestigious universities (sample 3). The gender gaps are estimated with equation 2. The regression is estimated with country-cohort-discipline and department fixed effects.

The figure suggests a "gender U" pattern. Publication gaps were small in periods with very low shares of female scientists. This can be referred to as the "Marie-Curie" period:

only exceptional women were hired and, on average, they published as much as men despite potential discrimination in the publication market. With increasing shares of women in the profession, gender gaps in publications turned negative. However, when the share of women increased beyond very low levels, the negative gender gaps in publications narrowed.

3.2 A Stylized Roy Model Of Gender Gaps in Hiring and Publications: "Gender U"

We outline a model along the lines of Roy (1951) to interpret the relationship between the share of women in academia and the observed gender gap in publications. The proposed 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 features allow the gender gap in publications to be explained by (a) *indirect* effects of selection and gender bias in the hiring market, and (b) the *direct* effect of gender bias in the publication 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. For example, certain academic high school tracks, many undergraduate programs, and most PhD programs did not accept women for a large part of the 20th century (e.g., Rossiter, 1982). In case of any such gender imbalance, we refer to the gender bias in hiring as Δ_0. We express selection in the hiring market in terms of s_0^W, the share of women among all hired academics. Second, for all hired academics, we observe publication outcomes at a later stage, denoted by 1, and we refer to any gender bias in publications as Δ_1. The model allows for the possibility that selection in the hiring stage 0 affects the observed publications in stage 1, so that the gender gap in publications may also be a function of Δ_0 and not only of Δ_1.

We first introduce a simplified version of the model that relies on the following assumptions:

(i) Δ_0 is not a function of s_0^W,

(ii) Δ_1 is not a function of s_0^W.

In Appendix C.2, we then present a generalized version of the model that relaxes assumption (ii) and other parametric assumptions (i.e., assumptions (iii)-(iv) below) and allows the gender gap in publications to be affected by the share of women in academia. In subsection 3.2, we present estimation results for both versions of the proposed Roy model.

Selection in the Hiring Market

Suppose that academic position i will be filled either by a woman W or by a man M. The latent value of hiring a woman is:

$$Y_{0i}^{W} = X_i^{W}\beta_0 + \epsilon_{0i}^{W}, \tag{3}$$

while that of hiring a man is:

$$Y_{0i}^{M} = X_i^{M}\beta_0 + \Delta_0 + \epsilon_{0i}^{M}, \tag{4}$$

where X_i^g, $g \in \{W, M\}$, are observable characteristics, Δ_0 a possible gender bias in hiring, and ϵ_{0i}^g the unobserved component of these latent valuations. As a result, academic position i will be filled by a woman if (3) is greater than (4):

$$\begin{aligned} Y_{0i} &= \left(X_i^W - X_i^M\right)\beta_0 - \Delta_0 + \left(\epsilon_{0i}^W - \epsilon_{0i}^M\right) \\ &= X_i\beta_0 - \Delta_0 + \epsilon_{0i} > 0 \end{aligned} \tag{5}$$

so that, keeping all other elements of the model fixed, when $\Delta_0 > 0$ women need to overcome the additional hurdle or gender bias Δ_0 to be hired. Assuming that ϵ_{0i} is distributed i.i.d. normal (assumption (iv) below), the probability that a woman is hired is:

$$\begin{aligned} s_0^W(X_i) &= \Pr[Y_{0i} > 0] = \Pr[\epsilon_{0i} > -X_i\beta_0 + \Delta_0] \\ &= \Phi\left(X_i\beta_0 - \Delta_0\right), \end{aligned} \tag{6}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal. It then follows that:

$$\Phi^{-1}\left(s_0^W(X_i)\right) = X_i\beta_0 - \Delta_0. \tag{7}$$

The differences in observable characteristics X_i could be SAT scores, college GPA, or the specialization of the undergraduate degree. Such data are not available on a world-wide scale over the 20th century and, if available could as well be affected by various sources of bias. Therefore, we assume $X_i = 0$, i.e., men and women are a priori equally qualified for the job, so that equation (7) reduces to $\Phi^{-1}\left(s_0^W\right) = -\Delta_0$, where s_0^W is the share of women among all academics. In this case, as we can directly measure s_0^W in the data, we compute Δ_0 without performing any estimation.[29]

[29] In settings with less comprehensive coverage but with more data on other observables X_i, one could estimate (β_0, Δ_0) by MLE using observations (Y_{0i}, X_i) and the probit model corresponding to equation (6).

Publication Market

Conditional on academic position i being filled by either a woman or a man, we have the following outcome equations at the publication stage:

$$\begin{aligned} Y_{1i}^W &= Z_i^W \beta_1 + \epsilon_{1i}^W & \text{if } Y_{0i} > 0 \\ Y_{1i}^M &= Z_i^M \beta_1 + \Delta_1 + \epsilon_{1i}^M & \text{if } Y_{0i} \leq 0, \end{aligned} \tag{8}$$

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. In words, if academic position i is filled by a woman, we observe the publication outcome of a woman, otherwise we observe the publication outcome of a man. Since for any i, we cannot observe the counterfactual publication outcome (i.e., the number of publications if position i had been filled by the other gender), equation (8) will be subject to endogenous selection on unobservables if the error terms in (5) and (8) are correlated: e.g., talented women, such as Marie Curie, were both more likely to get academic positions and to publish well once hired. Δ_1 is the gender bias in publications which may reflect gender imbalances in working conditions, biased editors or referees, preferences, or any other constraint that differentially affected publication outcomes of men and women.

In this simpler version of the model, we further make the two standard parametric assumptions (Heckman, 1979; Amemiya, 1984) of:

(iii) *Linearity*: $\epsilon_{1i}^g = \rho_g \epsilon_{0i} + \xi_i^g$, $g \in \{W, M\}$, with ξ_i^g independent of everything else in the model and with zero mean.

(iv) *Normality*: ϵ_{0i} is distributed i.i.d. normal.

As is well known, these are not necessary for identification but simplify estimation.[30] Parameter ρ_g measures the covariance between the unobserved component of selection in hiring, ϵ_{0i}, and the unobserved component of the publication outcome, ϵ_{1i}^g. Remember that $\epsilon_{0i} = \left(\epsilon_{0i}^W - \epsilon_{0i}^M\right)$, so that if $\rho_W > 0$, women that are more likely to get hired are also more likely to publish well. The same holds for men if $\rho_M < 0$.

Publication outcome conditional on gender

The expectation of Y_{1i}^W conditional on $Y_{0i} > 0$ is:

[30] As mentioned above, in Appendix C.2 we present a generalized version of the model that relaxes these assumptions as well as assumption (ii).

$$
\begin{aligned}
\mathbb{E}\left[Y_{1i}^W \middle| Z_i^W, Y_{0i} > 0\right] &= Z_i^W \beta_1 + \mathbb{E}\left[\rho_W \epsilon_{0i} + \xi_i^W \middle| \epsilon_{0i} > \Delta_0\right] \\
&= Z_i^W \beta_1 + \rho_W \, \lambda\left(-\Delta_0\right),
\end{aligned}
\tag{9}
$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio. Analogously, the expectation of Y_{1i}^M conditional on $Y_{0i} \leq 0$ is:

$$
\mathbb{E}\left[Y_{1i}^M \middle| Z_i^M, Y_{0i} \leq 0\right] = Z_i^M \beta_1 + \Delta_1 - \rho_M \, \lambda\left(\Delta_0\right). \tag{10}
$$

Combining equations (9) and (10) closes the model.[31] When the gender bias in publications Δ_1 increases, the conditional expectation of publications Y_{1i}^M increases. When the gender bias in hiring Δ_0 increases, i.e., only the best women are hired, the inverse Mills ratio $\lambda(\cdot)$ goes up, and the conditional expectation of publications Y_{1i}^W increases if $\rho_W > 0$. This holds independently of any gender bias Δ_1 in publications. In contrast, equation (10) indicates that when the gender bias in hiring Δ_0 increases, because $\lambda(\cdot)$ approaches zero, men's conditional expectation of Y_{1i}^M may be unaffected, even if $\rho_M \neq 0$. Note that, however, when $\rho_M > 0$ relatively less productive men become academics, $-\rho_M \, \lambda(\Delta_0) < 0$, a phenomenon which will tend to attenuate the observed gender gap in publications.

The model highlights how selection and gender biases in hiring can indirectly affect estimated gender gaps in publications (in addition to any direct gender bias in publications). Note that in this model the inverse Mills ratio does not depend on Z_i^W and Z_i^M but only on the observed share of women (information otherwise not used in the publication outcome equation), which serves the purpose of an exclusion restriction.

Estimation Results

Based on the model, we specify a regression that directly estimates all the components of the gender gap in publications, the direct gender bias in publications $-\Delta_1$, and the indirect effect as a function of the share of female scientists. In practice, we estimate (9) and (10) on the basis of sample 3 (scientists, prestigious universities, 1900-2000) with a regression similar to (1) that also includes interactions with the inverse Mills ratios computed using the female shares in the respective country and cohort:

[31] If data on X_i were available, the inverse Mills ratios would instead be $\lambda(X_i\beta_0 - \Delta_0)$ in (9) and $\lambda(\Delta_0 - X_i\beta_0)$ in (10).

$$\begin{aligned}
\text{Pub}_{idt} &= \gamma + \text{Female}_{idt} \times 1\left[t(i) = 1900/38\right] \times \left[\rho_1^W \lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \Delta_{11} + \rho_1^M \lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)\right] \\
&+ \text{Female}_{idt} \times 1\left[t(i) = 1956/69\right] \times \left[\rho_2^W \lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \Delta_{12} + \rho_2^M \lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)\right] \\
&+ \text{Female}_{idt} \times 1\left[t(i) = 2000\right] \times \left[\rho_3^W \lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \Delta_{13} + \rho_3^M \lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)\right] \\
&+ \text{Experience}_{idt}\gamma_{\exp} + \text{Fixed Effects} + \varepsilon_{idt},
\end{aligned} \tag{11}$$

where Pub_{idt} measures the standardized number of papers published by scientist i in cohort $t(i)$ and department d, $s_{0\ell(i)}^W$ is the share of female academics in scientist i's country-cohort $\ell(i)$ (e.g., USA in 2000), and the fixed effects are at the level of the cohort $\times$ discipline $\times$ country and department.[32] The regression results are reported in Appendix Table C.2. The results show a negative gender bias ($-\widehat{\Delta}_{1p}, p = 1, 2, 3$) in publications, suggesting that women faced additional hurdles to publish throughout the 20th century. Such hurdles could come from biased editors or referees (e.g., Card et al., 2020) or from women facing other constraints to publish, e.g., worse working conditions or additional childcare (e.g., Moser and Kim, 2022).

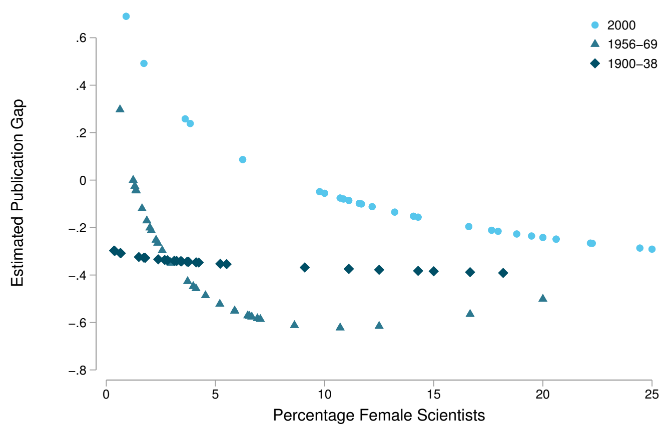
Furthermore, the estimates of $\widehat{\rho}_p^W$, $\widehat{\rho}_p^M, p = 1, 2, 3$ are positive in all time periods. This indicates that the persistent gap at the hiring stage Δ_0 allowed only the most productive women and relatively less productive men to enter academia throughout the 20th century, highlighting that women were positively selected into academia throughout the 20th century. A finding that is similar to Ashraf et al. (2022), who document positive selection of women into working for a large multinational firm in the period 2015-2019.

In Figure 7, panel (a), we summarize the OLS estimation results of regression (11) by plotting, separately for each period p, the predicted gender gap in standardized publications $\widehat{\rho}_p^W \lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \widehat{\Delta}_{1p} + \widehat{\rho}_p^M \lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)$ as a function of $s_{0\ell}^W$ in each country. Each dot in the figure represents the estimated gender gap in publications for each country-cohort as a function of the share of women in that country-cohort (separately for three periods: 1900-1938, 1956-1969, and 2000). In Figure 7, panel (b), we plot the analogous predictions for the version of the Roy model described in Appendix C.2 that approximates the gender gap functions by second-degree polynomials of $s_{0\ell}^W$.

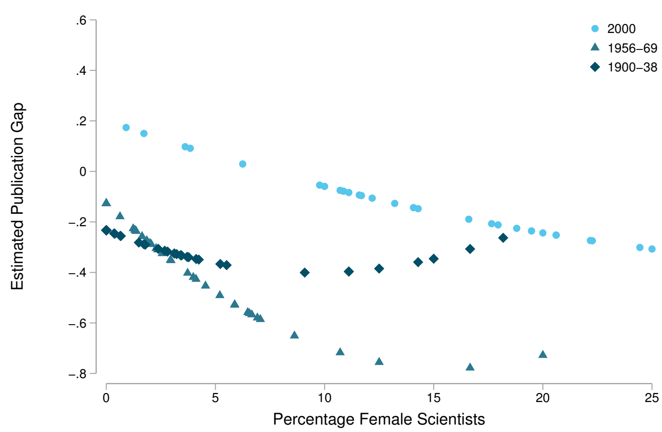
[32] For cohort-country combinations with $s_{0\ell}^W$ equal to zero, $\Delta_0 = -\Phi^{-1}\left(s_{0\ell}^W\right)$ is not defined. As a consequence, the regressions exclude observations from cohort-country combinations that do not include any women. In contrast, the Roy model described in Appendix C.2 is well defined for cohort-country combinations with $s_{0\ell}^W$ equal to zero. However, for comparability with regression (11), we also exclude observations corresponding to cohort-country combinations that do not include any woman from the estimation of regression (C.8). Estimation results of regression (C.8) are however robust to the inclusion of these observations.

Figure 7: Gender Gaps in Publications and the Share of Female Academics

(a) Inverse Mills Ratios



(b) Polynomial Approximations



Notes: The Figure plots the estimated gender gap functions in standardized publications as a function of the percentage of female scientists by country and cohort. Each dot in the figure represents the estimated gender gap in publications for a country-cohort as a function of the share of women in that country-cohort, e.g. United States - 1938 or United States - 2000. Panel a plots estimated gender gap functions from regression (11), while panel b from regression (C.8) that approximates the gender gap functions by second-degree polynomials of $s_{0\ell}^{W}$ (see Appendix C.2 for details). Appendix table C.2 details the estimation results of regressions (11) and (C.8).

The figures suggest that publication gaps are small or strongly positive in countries with very low shares of female scientists: the "Marie-Curie" period. However, with increasing shares of women in the profession, gender gaps in publications become more negative. When the share of women increases beyond very low levels, the negative gender gap in publications narrows for most time periods. The results indicate that a Roy model can rationalize the "gender U" pattern from Figure 6, stressing that gender biases in hiring have indirect repercussions on the observed output of women. This finding crucially relies on our ability to observe gender gaps in both hiring and publications for the same set of academics.

4 Gender Gaps in Citations: Controlling for Predicted Citations

In the previous section, we have shown that women published fewer papers than men. This section explores if papers written by women were also cited less. To this aim, we turn to a paper-level analysis of citation gaps.

4.1 A Novel Procedure to Predict Citations

A key challenge when estimating citation gaps is that men and women may work on different topics which may affect citations. To overcome this issue, we develop a novel supervised machine learning approach that controls for the citation potential of papers. It predicts the citations of each paper, as if the paper had been written by men. The approach uses words from papers' titles and uncovers complex relationships between research topics and citations (see Appendix D for further details).[33] A similar approach could be used to study performance or pay gaps between groups if one had access to highly granular data (e.g., detailed occupational task descriptions, the full text of job advertisements, or performance reviews).

For the machine learning approach, the training sample solely consists of papers published by men. This prevents the model from internalizing biases against papers published by women (see Barocas and Selbst (2016) for an overview of AI biases). In preparation for the machine learning, we filter all non-alphanumerical characters from papers' titles, remove common words (stopwords, e.g., "the"), and stem the words in the titles. 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 $\boldsymbol{X}$ 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 the following cost function:

[33] A pre-trained model of predicted citations is available at carloschwarz.eu/programming/. The model allows to predict the log number of citations from the titles of papers. We also provide a Python and Stata wrapper (see also Schwarz, 2022).

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

$$C = \min_{(\omega_j)_{j=1}^{W}} \left\{ \sum_{p=1}^{N} \left(y_p - \sum_{j=1}^{W} \omega_j \cdot x_{pj} \right)^2 + \lambda \sum_{j=1}^{W} \omega_j^2 \right\}, \tag{12}$$

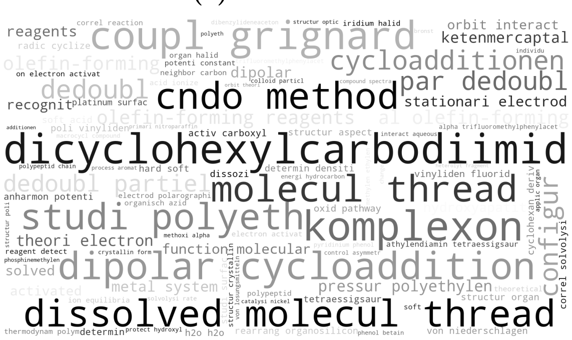
where y_p are the total citations of paper p (standardized by country, discipline, and cohort). To avoid results to be driven by a few outliers, we winsorize citations at the 99.9th percentile (by discipline, and cohort).[35] The main explanatory variables are the W variables indicating the unigrams and bigrams that correspond to the respective entries of the paper-1,2-gram-matrix $\boldsymbol{X}$. We additionally include a full set of indicators for the number of words in the title of each paper. We train separate models for each discipline in each of the seven cohorts. For each discipline and cohort, we choose the optimal normalization strength λ using 10-fold cross validation. The algorithm predicts citations $\hat{y}_p$ for each paper p.

Figure 8: Words that Predict High Citations in Chemistry over Time

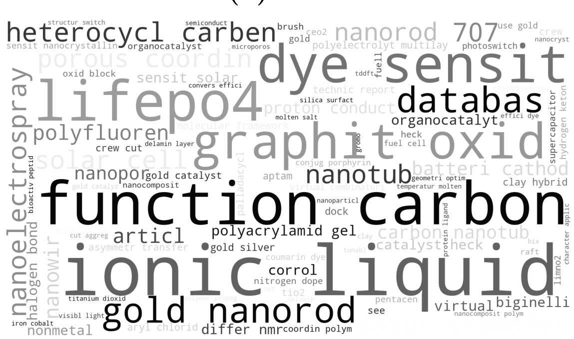
(a) 1900-1914

(b) 1925-1938

(c) 1965-1969



(d) 2000



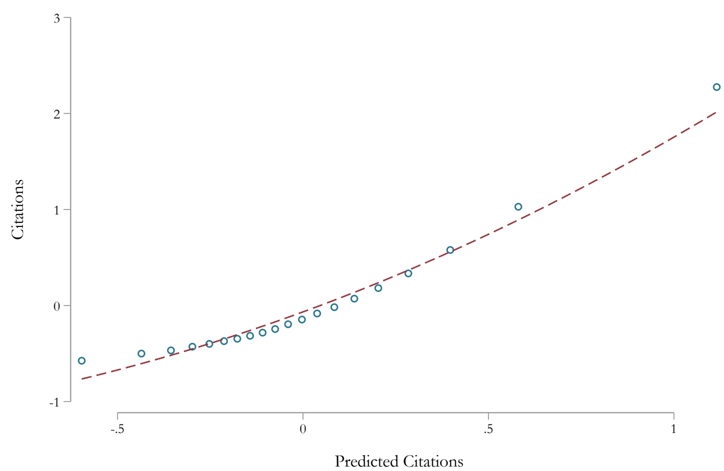
Notes: The Figure shows the unigrams and bigrams that predict the highest citations in chemistry for the indicated cohorts. The n-grams are identified with an L2-regularized regression model (ridge regression) that uses unigrams and bigrams of the title as inputs, see section 4.1 for details. A very small fraction of words in the titles are not translated to 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 8 summarizes the unigrams and bigrams that predict high citations for chemistry and how they

[35] The results are very similar if we do not winsorize the citation counts (see Appendix Table D.1)

evolved between 1900 and 2000.[36] For example, for the 1900 and 1914 cohorts, the classifier detects the names of sugar molecules (e.g., "glucose", "dextrose", "galactose"). The understanding of the structure of sugar molecules was developed by German chemist and 1902 Nobel Prize winner Emil Fischer in the 1890s. Other highly cited n-grams in the same cohort are "triketohydrindane" and "triketohydrindane hydrate" (alternative names for the compound Ninhydrin), a compound discovered by Siegfried Ruhemann in 1910. For the 1925 and 1938 cohorts, the classifier detects the words "elastic" and "thread," which refer to the discoveries of the first polymers, i.e., Nylon by the DuPont chemist Wallace Carothers in 1935 and Perlon by the I.G. Farben chemist Paul Schlack in 1938. For the 1956 and 1969 cohorts, the classifier detects "CNDO method," an abbreviation for "Complete Neglect of Differential Overlap." CNDO is one of the first methods in quantum chemistry and was developed in the 1960s by chemistry Nobel laureate John Pople. Lastly, for the 2000 cohort, the classifier detects "LiFePo4," which is the chemical formula for lithium iron phosphate — a cathode material that can be used for batteries and which was discovered in 1996. Another interesting stem for the 2000 cohort is "organocatalyt", showing the importance of the field of organocatalysis for which Benjamin List and David Macmillan were awarded the Nobel Prize in 2021.

Figure 9: Predicted and Actual Citations



Notes: The Figure shows the relationship between actual and predicted ciations. Actual citations is the count of citations of each paper which we standardize at the cohort-country-discipline 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 9 visualizes the strong positive

[36]To save space, we show word-clouds combining multiple cohorts. However, for the regression results reported below, we implement the prediction at the cohort by cohort.

relationship between predicted and actual citations ($R^2 = 0.35$).[37] As the figure suggests, the R^2 increases further when we include a second-order polynomial of the predicted citations measure. In contrast, higher-order polynomials do not lead to any further increases in the R^2. We therefore control for the first and second-degree polynomial of the value of predicted citations in our baseline regressions.[38]

4.2 Paper-Level Citation Gaps — Controlling for Predicted Citations

In the following analysis, we estimate the citation gaps at the paper level, depending on whether papers were published by men or women. Importantly, we add our novel measure of predicted citations as a regressor to control for the fact that women may be working on topics with less citation potential than men.[39] We estimate the following regression:

$$\begin{aligned}
\text{Citations}_{pt} &= \gamma_1 + \gamma_2 \text{Female}_{pt} \times 1[t(i) = 1900/14] + \gamma_3 \text{Female}_{pt} \times 1[t(i) = 1925/38] \\
&+ \gamma_4 \text{Female}_{pt} \times 1[t(i) = 1956/69] + \gamma_5 \text{Female}_{pt} \times 1[t(i) = 2000] \\
&+ \gamma_6 \widehat{\text{Citations}}_{pt} + \text{Fixed Effects} + \xi_{pt}
\end{aligned} \tag{13}$$

The dependent variable is the number of (standardized) citations of paper p. The main explanatory variables are the interactions of the indicator variable Female_{pt} with indicators for four different time periods: pre-WW1 (1900 and 1914 cohorts), interwar (1925 and 1938), post-WW2 (1956 and 1969), and modern (2000). Female_{pt} is an indicator that is equal to one if a paper has at least one female author.[40] Importantly, we control for the predicted citations of paper p by including $\widehat{\text{Citations}}_p$, the first and second-degree polynomial of the value of

[37] Note 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 for details).

[38] In Appendix Table D.1, we show that the results are almost identical if we alternatively control for predicted citations either linearly or non-parametrically.

[39] In a recent paper, Koffi (2021) proposes a different methodology to estimate whether papers by female authors are under-cited. Her methodology uses text similarity to identify existing papers that the focal paper should have cited. In contrast, our methodology controls for differences in the citation potential of papers. A key advantage of our methodology is that it can be used when female shares are very low and, hence, papers by women are unlikely to be among the most similar papers. Furthermore, our methodology can also be used if the number of papers is large, due to its larger computational efficiency. The methodology by Koffi (2021) requires the calculation of the pairwise similarity between all papers in the same field. Hence, the number of necessary calculations grows quadratically in the number of papers.

[40] The results are very similar if we instead use the share of female authors. For most papers, the share of female authors is either 0 or 1 (see Appendix Figure D.1). Our preferred specifications use an indicator variable because we only know the gender of authors who can be linked to the faculty rosters, but we do not know the gender of the other authors (e.g., graduate students). Hence, a variable measuring the female share suffers from measurement error, while the female indicator does not.

predicted citations for paper p. We also control for additional fixed effects. In the baseline specification, we control for the three-way interaction of cohort, discipline, and country fixed effects.[41] We cluster standard errors at the discipline-country level. To account for the fact that the predicted citation control is estimated, we additionally report cluster-bootstrap standard errors in square brackets.

Papers published by the 1900 and 1914 cohorts of female scientists received 0.28 standard deviations (s.d.) fewer citations than papers published by male scientists in the sample containing all universities (sample 2). Papers published by the 1925 and 1938 cohorts of female scientists received 0.16 s.d. fewer citations, while papers published by the 1956 and 1969 cohorts of female scientists received 0.1 s.d. fewer citations (Table 2, sample 2, column 1). When we control for predicted citations to address the issue that women may have been working on topics that attract fewer citations, the estimated gaps fall slightly but remain significant (Table 4, sample 2, column 2). The results are similar if we compare papers published by scientists in the same department, or even department by cohort (e.g., Harvard in 1969). The results suggest that papers by female scientists received fewer citations but that the gender gap in citations declined by around 64 percent from 1900 until 1969.

The estimated citation gaps are similar for papers published by authors from prestigious universities (sample 3) and exhibit a similar decline over time. For the 2000 cohort, the estimated citation gaps are very close to 0 and precisely estimated, indicating that by the year 2000, men and women receive the same number of citations for similar papers.

In Table 5, we explore further reasons that could explain why papers by female authors may receive fewer citations. After controlling for the detailed topics of papers, there are at least three additional reasons that may explain this finding. First, women may have fewer opportunities to write papers with co-authors (production effect). This may translate into fewer citations because co-authored papers, on average, receive more citations (e.g., Wuchty, Jones, and Uzzi, 2007). Second, women may publish their papers in lower-ranked journals because of biased editors or referees (publication effect). Such an effect has been shown for economics papers (Card et al., 2022). Third, papers by female authors may receive fewer citations because of biases in the citation market. Biases could arise, for example, because women have fewer opportunities to present their work or because of discrimination.

[41] For these results, the country and department fixed effects are defined at the paper level. To account for multi-university research teams, we include separate fixed effects for any combination of departments. For example, a paper co-authored by scientists from Harvard and MIT has a separate fixed effect from papers authored by Harvard scientists only.

Table 4: Gender Gaps in Paper-Level Citations: Controlling for Predicted Citations

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Standardized Citations					
Sample 2: All Universities 1900-1969						
Female-authored Paper (1900/14)	-0.284***	-0.268**	-0.163	-0.252*	-0.182***	-0.313***
	(0.097)	(0.106)	(0.112)	(0.133)	(0.033)	(0.017)
		[0.186]		[0.233]		[0.211]
Female-authored Paper (1925/38)	-0.162***	-0.141***	-0.195***	-0.124*	-0.176***	-0.118**
	(0.029)	(0.052)	(0.046)	(0.070)	(0.047)	(0.057)
		[0.074]		[0.098]		[0.086]
Female-authored Paper (1956/69)	-0.099***	-0.080***	-0.096***	-0.102***	-0.082**	-0.112***
	(0.034)	(0.025)	(0.031)	(0.033)	(0.035)	(0.036)
		[0.041]		[0.044]		[0.049]
Observations	155,264	155,264	155,264	155,264	155,264	155,264
R^2	0.002	0.443	0.083	0.473	0.111	0.483
Sample 3: Prestigious Universities 1900-2000						
Female-authored Paper (1900/14)	-0.366***	-0.341***	-0.229***	-0.316***	-0.209***	-0.282***
	(0.079)	(0.048)	(0.078)	(0.019)	(0.018)	(0.004)
		[0.13]		[0.143]		[0.142]
Female-authored Paper (1925/38)	-0.175***	-0.144***	-0.235***	-0.146***	-0.167***	-0.129**
	(0.026)	(0.050)	(0.036)	(0.041)	(0.049)	(0.055)
		[0.078]		[0.073]		[0.076]
Female-authored Paper (1956/69)	-0.108**	-0.106***	-0.089	-0.108**	-0.071	-0.127***
	(0.043)	(0.031)	(0.060)	(0.049)	(0.064)	(0.048)
		[0.055]		[0.062]		[0.06]
Female-authored Paper (2000)	-0.022	-0.016	-0.023	-0.021	-0.027	-0.021
	(0.014)	(0.014)	(0.018)	(0.018)	(0.019)	(0.018)
		[0.017]		[0.019]		[0.019]
Observations	274,447	274,447	274,447	274,447	274,447	274,447
R^2	0.002	0.377	0.063	0.406	0.082	0.412
Predicted Citiation Control		Yes		Yes		Yes
Country × Discipline × Cohort FE	Yes	Yes	Yes	Yes		
Department FE			Yes	Yes		
Cohort × Department FE					Yes	Yes

Notes: The Table shows gender gaps in citations per paper. Results are estimated at the paper-level. The dependent variable is the citation count which we standardize at the cohort-country-discipline level. The main explanatory variable and indicator variable which is equal to 1 if the paper has at least 1 female author. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for the first and second degree polynomial of the predicted citation variable. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model (ridge regression), see section 4.1 for details. Standard errors are clustered at the discipline-country level. We additionally report boostrap standard errors in square brackets. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

Table 5: Gender Gaps in Paper-Level Citations: Controlling for Author and Journal Fixed Effects

	Baseline		Author Nr. FE		Journal FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Standardized Citations					
Sample 2: All Universities 1900-1969						
Female-authored Paper (1900/14)	-0.284***	-0.268**	-0.284***	-0.268**	-0.337***	-0.288***
	(0.097)	(0.106)	(0.092)	(0.104)	(0.082)	(0.094)
		[0.186]		[0.184]		[0.177]
Female-authored Paper (1925/38)	-0.162***	-0.141***	-0.165***	-0.143***	-0.157***	-0.141**
	(0.029)	(0.052)	(0.029)	(0.051)	(0.045)	(0.055)
		[0.074]		[0.074]		[0.074]
Female-authored Paper (1956/69)	-0.099***	-0.080***	-0.099***	-0.081***	-0.070***	-0.057**
	(0.034)	(0.025)	(0.035)	(0.025)	(0.026)	(0.025)
		[0.041]		[0.04]		[0.042]
Observations	155,264	155,264	155,264	155,264	155,264	155,264
R^2	0.002	0.443	0.003	0.443	0.064	0.461
Sample 3: Prestigious Universities 1900-2000						
Female-authored Paper (1900/14)	-0.366***	-0.341***	-0.359***	-0.337***	-0.392***	-0.355***
	(0.079)	(0.048)	(0.077)	(0.047)	(0.086)	(0.051)
		[0.13]		[0.13]		[0.129]
Female-authored Paper (1925/38)	-0.175***	-0.144***	-0.178***	-0.146***	-0.134***	-0.125**
	(0.026)	(0.050)	(0.024)	(0.050)	(0.043)	(0.054)
		[0.078]		[0.077]		[0.078]
Female-authored Paper (1956/69)	-0.108**	-0.106***	-0.110**	-0.107***	-0.074**	-0.079**
	(0.043)	(0.031)	(0.044)	(0.031)	(0.033)	(0.037)
		[0.055]		[0.054]		[0.058]
Female-authored Paper (2000)	-0.022	-0.016	-0.036**	-0.022	-0.014	-0.014
	(0.014)	(0.014)	(0.015)	(0.014)	(0.012)	(0.012)
		[0.017]		[0.017]		[0.015]
Observations	274,447	274,447	274,447	274,447	274,447	274,447
R^2	0.002	0.377	0.008	0.379	0.118	0.425
Predicted Citations Control		Yes		Yes		Yes
Country × Discipline × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Nr. Authors FE			Yes	Yes		
Journal FE					Yes	Yes

Notes: The Table shows gender gaps in citations. Results are estimated at the paper level. The dependent variable is the winsorized citation count which we standardize at the cohort-country-discipline level. The main explanatory variable and indicator variable which is equal to 1 if the paper has at least 1 female author. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for the first and second degree polynomial of the predicted citation variable or for 1000 dummies for the quantiles of the predicted citation distribution. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model. Standard errors are clustered at the discipline-country level. We additionally report boostrap standard errors in square brackets. Significance levels: ∗∗∗ $p<0.01$, ∗∗ $p<0.05$, and ∗ $p<0.1$.

We explore the first possible explanation by controlling for the number of authors of each paper (i.e., a fixed effect if the paper has one author, another fixed effect if the paper has two authors, and so on). Interestingly, controlling for the number of authors does not affect gender gaps in citations (Table 5, columns 3-4). Next, we explore the publication effect by including a full set of journal fixed effects (columns 5-6). The inclusion of these has little impact on the magnitude of the estimates. Of course, both the number of authors and the journals in which papers are published are potentially endogenous. Therefore these results should be interpreted as a decomposition of the citation gaps. Notably, even after controlling for the number of co-authors, and the journal, female-authored papers received significantly fewer citations until the late 1960s. This suggests that biases in the citation market may have reduced citations of female academics.

In Appendix Table D.1, we show that the results are robust to using alternative functional forms for the predicted citation control (columns 2 and 3). We also address the concern that the model absorbs more variation for male-authored papers, because the model is trained on the papers written by men. For this check, we train an "out of sample" version of our model. This version is trained on a 50% hold-out sample for each cohort and subject. This alternative approach comes at the cost of excluding papers from the outcome regression. Citation gaps using the "out of sample" model are slightly larger in absolute magnitude than those using the baseline model (column 4). The results are also very similar if we do not winsorize the citations (column 5). Lastly, we show that results are robust to using citation counts, instead of standardized citations, as the dependent variable (column 6).

5 Gender Gaps in Promotions

In the last section, we investigate gender gaps in promotions. We focus on the sample of academics who were not already full professors when they entered the dataset in cohort t.[42] We then analyze whether they get promoted to full professor by cohort $t+1$ (see Appendix A.1.1 for more information on the coding of promotions). Promotions to full professor are particularly important because full professors have unique privileges and high job security and salaries. We estimate the following regression:

$$\begin{aligned} \text{Promotion Full Prof}_{idt} &= \pi_1 + \pi_2 \text{Female}_{idt} \times 1[t(i) = 1914] + \pi_3 \text{Female}_{idt} \times 1[t(i) = 1925/38] \\ &+ \pi_4 \text{Female}_{idt} \times 1[t(i) = 1956/69] + \pi_5 \text{Female}_{idt} \times 1[t(i) = 2000] \\ &+ \text{Fixed Effects} + \upsilon_{idt}. \end{aligned} \tag{14}$$

[42]This restriction results in a smaller sample because academics who enter the data as full professors cannot be included in the analysis. Furthermore, all academics who enter the data in the last cohort (independently of their rank) cannot be included in the analysis. Lastly, we can only analyze promotions for academics who we observe in at least two cohorts of the data. The probability that we observe women in at least two cohorts is smaller than for men. As a result, the gender gaps in promotions are likely lower bound estimates.

The dependent variable Promotion Full Prof_{idt} is an indicator for whether academic i who entered the data in department d and cohort t was promoted to full professor by cohort $t+1$ (in any department). The main explanatory variables are the interactions of the indicator variable Female_{idt} with indicators for the four different time periods. As before, the regressions include fixed effects that control for country, discipline, and over time differences, e.g., for the fact that in certain disciplines, countries, and or time periods, there may be more openings of full professor positions.
In all universities and disciplines (sample 1), women who started their careers in the 1900 cohort were, on average, 17 percentage points less likely to be promoted to full professor than men by 1914 (Table 2, sample 1, column 1, significant at the 5 percent level). Because the probability of promotion to full professor was around 64 percent, women were about 27 percent less likely to be promoted. Women who started their careers in the 1925 to 1938 cohorts and in the 1956 to 1969 cohorts, women were around 22 percentage points (or 34 percent) less likely to be promoted to full professor by cohort $t+1$. 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 department and cohort, even though the estimated gap is no longer significant for the 1900 and 1914 cohorts (Table 6, sample 1, column 3).

In the scientist sample (mathematics, chemistry, and biochemistry) of all universities (sample 2), women who started their careers in the 1900 cohort were around 40 percentage points less likely to be promoted by 1914. Women who started their careers in the 1925 to 1938 cohorts and in the 1956 to 1969 cohorts were 11 percentage points (not significantly different from 0) and 29 percentage points less likely to be promoted to full professor by cohort $t+1$ (Table 6, sample 2, column 1). The promotion gaps are relatively similar if we condition on more stringent fixed effects.

The results are also similar if we estimate gender gaps in promotions for scientists in prestigious universities (sample 3). In this sample we can extend the time horizon until 2000. By that point, the gender gap in promotions has disappeared. It is important to note that the number of women in the sciences was low, especially during the first part of the 20th century. As promotions can only be observed for scientists who enter the data as assistant or associate professors in cohort t some estimates lack power, even more so if we compare men and women who start their careers in the same department.

Table 6: Gender Gaps in Promotions

	(1)	(2)	(3)	(4)
Dependent Variable:		Promotion Indicator		
Sample 1: All Universities, all disciplines, 1900-1969				
Female (1914)	-0.169**	-0.224**	-0.120	
	(0.084)	(0.104)	(0.260)	
Female (1925/38)	-0.223***	-0.245***	-0.243***	
	(0.033)	(0.036)	(0.047)	
Female (1956/69)	-0.217***	-0.196***	-0.192***	
	(0.025)	(0.025)	(0.028)	
Observations	31,842	31,842	31,842	
R-squared	0.229	0.453	0.562	
Sample 2: All Universities, sciences, 1900-1969				
Female (1914)	-0.396**	-0.430***	0.000	0.000
	(0.177)	(0.041)	(0.000)	(0.000)
Female (1925/38)	-0.110	-0.179*	-0.225**	-0.219**
	(0.111)	(0.106)	(0.091)	(0.090)
Female (1956/69)	-0.293***	-0.312***	-0.351***	-0.334***
	(0.043)	(0.100)	(0.075)	(0.082)
Std. Publications				0.057***
				(0.019)
Std. Citations				0.009
				(0.013)
Observations	3,182	3,182	3,182	3,182
R-squared	0.197	0.459	0.646	0.657
Sample 3: Prestigious Universities, sciences, 1900-2000				
Female (1914)	-0.382***	-0.380***	0.000	0.000
	(0.000)	(0.018)	(0.000)	(0.000)
Female (1925/38)	-0.068	-0.224*	-0.189*	-0.182
	(0.108)	(0.115)	(0.113)	(0.111)
Female (1956/69)	-0.234***	-0.258**	-0.313***	-0.293**
	(0.059)	(0.128)	(0.110)	(0.117)
Female (2000)	0.058	0.046*	0.000	-0.007
	(0.040)	(0.025)	(0.000)	(0.019)
Std. Publications				0.044**
				(0.020)
Std. Citations				0.020
				(0.013)
Observations	2,441	2,441	2,441	2,441
R-squared	0.257	0.454	0.669	0.680
Cohort × Discipline × Country FE	Yes	Yes		
Department FE		Yes		
Cohort × Department 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 covering the period 1900-1969. Sample 2 includes scientists (mathematicians, chemists, and biochemists) in all universities covering the period 1900-1969. Sample 3 includes scientists in prestigious universities covering the period 1900-2000. The dependent variable is an indicator that equals 1 if an academic who entered the dataset in cohort t was promoted to full professor by cohort $t+1$. The main explanatory variable is an indicator that equals 1 if the academic is a woman. The regressions also control for various fixed effects, as indicated in the table. For samples 2-3, the regressions also control for the publication and citation record of the scientist. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

For scientists (samples 2 and 3), we can also control for their publication and citation records. While publications have a significant effect on promotions, controlling for these records hardly affects the gender gap in the promotion to full professor.[43] Strikingly, the unexplained gender gap in promotions to full professor is larger than the effect of a four to five standard deviations worse publication record. This is remarkable because the true quality of women conditional on the same number of publications and citations should be, if anything, higher in the presence of discrimination and other biases in the publication market.

6 Conclusion

Leveraging new worldwide data on academics, this paper sheds light on the evolution of gender gaps in academia. From our analysis, four results stand out. First, only one percent of academics were women in 1900, and the share of women increased slowly until 1969 but faster in the last three decades of the 20th century. However, by the end of the 20th century women remain underrepresented in the sciences in prestigious universities. Second, gender gaps in publications evolved according to a U-shape over time: publication gaps were close to 0 in times with very low shares of women, then widened until 1970, only to narrow again until 2000. Third, papers by female authors received fewer citations. These citation gaps are not explained by differences in the research topics women worked on. Lastly, female academics were less likely to be promoted, even compared to male peers within the same department, cohort, and with the same publication and citation record; indicating pervasive unequal opportunities in academia throughout the 20th century.

Together, these patterns depict a new and rich portrait of women's entry in academia. Our findings highlight fruitful directions for future research and reveal the important role that countries, universities, and disciplines played in the participation of women in academia.

[43] In additional results, we control more flexibly for publications and citations by including indicators for each percentile of the publication and citation distribution. The results are very similar to the ones with the linear publication and citation controls.

References

1930. "Index Generalis: Annuaire Général des Universités, Grandes Écoles, Académies, Archives, Bibliothèques, Institute scientifiques, Jardins botaniques et zoologiques, Musées, Observatoires, Stés Savantes." Tech. rep.

Acemoglu, Daron, David H. Autor, and David Lyle. 2004. "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury." *Journal of Political Economy* 112 (3):497–551.

Albanesi, Stefania, Claudia Olivetti, and María José Prados. 2015. "Gender and Dynamic Agency: Theory and Evidence on the Compensation of Top Executives." *Research in Labor Economics* 42:1–59.

Albrecht, James, Anders Björklund, and Susan Vroman. 2003. "Is there a glass ceiling in Sweden?" *Journal of Labor Economics* 21 (1):145–177.

Altonji, Joseph G. and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." *Handbook of Labor Economics* 3:3143–3259.

Amemiya, Takeshi. 1984. "Tobit models: A survey." *Journal of Econometrics* 24 (1-2):3–61.

Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Víctor Quintas-Martínez. 2022. "Gender Roles and the Misallocation of Labour Across Countries." *mimeo* .

Azmat, Ghazala and Rosa Ferrer. 2017. "Gender gaps in performance: Evidence from young lawyers." *Journal of Political Economy* 125 (5):1306–1355.

Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender differences in accepting and receiving requests for tasks with low promotability." *American Economic Review* 107 (3):714–47.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the gender composition of scientific committees matter?" *American Economic Review* 107 (4):1207–38.

Barocas, Solon and Andrew D. Selbst. 2016. "Big data's disparate impact." *Calif. L. Rev.* 104:671.

Becker, Sascha O., Volker Lindenthal, Sharun Mukand, and Fabian Waldinger. 2021. "Scholars at Risk: Professional Networks and High-Skilled Emigration from Nazi Germany." .

Beidas, Rinad S., Peggy A. Hannon, Aimee S. James, and Karen M. Emmons. 2022. "Advancing gender equity in the academy." *Science Advances* 8 (13):eabq0430.

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

Bertrand, Marianne. 2011. "New perspectives on gender." In *Handbook of Labor Economics*, vol. 4. Elsevier, 1543–1590.

Bertrand, Marianne, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway." *The Review of Economic Studies* 86 (1):191–239.

Bertrand, Marianne and Esther Duflo. 2017. "Field experiments on discrimination." *Handbook of Economic Field Experiments* 1:309–393.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics* 2 (3):228–55.

Bertrand, Marianne and Kevin F. Hallock. 2001. “The gender gap in top corporate jobs.” *ILR Review* 55 (1):3–21.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2008. “Gender wage disparities among the highly educated.” *Journal of Human Resources* 43 (3):630–659.

Blau, Francine D. and Lawrence M. Kahn. 2017. “The gender wage gap: Extent, trends, and explanations.” *Journal of Economic Literature* 55 (3):789–865.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2020. “Are referees and editors in economics gender neutral?” *The Quarterly Journal of Economics* 135 (1):269–327.

———. 2022. “Gender Differences in Peer Recognition by Economists.” *Econometrica* .

Curie, Eve. 1938. *Madame Curie.* Gallimard Paris.

De la Croix, David, Frédéric Docquier, Alice Fabre, and Robert Stelter. 2020. “The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000-1800).” .

Gayle, George-Levi, Limor Golan, and Robert A. Miller. 2012. “Gender differences in executive compensation and job mobility.” *Journal of Labor Economics* 30 (4):829–872.

Goldin, Claudia. 1989. “Life-cycle labor-force participation of married women: Historical evidence and implications.” *Journal of Labor Economics* 7 (1):20–47.

———. 1990. *Understanding the gender gap: An economic history of American women.* National Bureau of Economic Research.

———. 1991. “The role of World War II in the rise of women’s employment.” *The American Economic Review* :741–756.

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

Heckman, James J. 1979. “Sample selection bias as a specification error.” *Econometrica* :153–161.

Hengel, Erin. 2020. “Publishing while female (summary).” .

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

Huang, Junming, Alexander J. Gates, Roberta Sinatra, and Albert-László 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, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. “Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science.” *The Quarterly Journal of Economics* 133 (2):927–991.

Jones, Charles I. 1995a. “R & D-based Models of Economic Growth.” *Journal of Political Economy* 103 (4):759–784.

Koffi, Marlène. 2021. “Innovative ideas and gender inequality.” Tech. rep., Working Paper Series.

Kricheldorf, Hans R. 2014. *Erkenntnisse und Irrtümer in Medizin und Naturwissenschaften.* Springer-Verlag.

McGrayne, Sharon Bertsch. 1998. *Nobel Prize Women in Science: their Lives, Struggles, and Momentous Discoveries.* Birch Lane Press.

Moser, Petra and Scott Kim. 2022. “Women in Science. Lessons from the Baby Boom.” .

Mulligan, Casey B. and Yona Rubinstein. 2008. "Selection, investment, and women's relative wages over time." *The Quarterly Journal of Economics* 123 (3):1061–1110.

OECD. 2021. "Inventors." URL https://data.oecd.org/entrepreneur/inventors.htm.

Olivetti, Claudia and Barbara Petrongolo. 2016. "The evolution of gender gaps in industrialized countries." *Annual Review of Economics* 8:405–434.

Ranking, Shanghai. 2020. "Academic Ranking of World Universities 2020." URL http://www.shanghairanking.com/index.html.

Romanowicz, Barbara. 2019. "Equality of opportunity in the European Research Council Competitions." In *EGU General Assembly Conference Abstracts*. 10306.

Romer, P. M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5):S71–S102.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5):1002–1037.

Ross, Matthew, Britta Glennon, Raviv Murciano-Goroff, Enrico Berkes, Bruce Weinberg, and Julia Lane. 2022. "Women are Credited Less in Science than are Men." *Nature* .

Rossiter, Margaret W. 1982. *Women Scientists in America: Struggles and Strategies to 1940*, vol. 1. JHU Press.

Roussille, Nina. 2022. "The central role of the ask gap in gender pay inequality." *mimeo* .

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers* 3 (2):135–146.

Santamaría, Lucía and Helena Mihaljević. 2018. "Comparison and Benchmark of Name-to-Gender Inference Services." *PeerJ Computer Science* 4:e156.

Sarsons, Heather. 2017. "Recognition for group work: Gender differences in academia." *American Economic Review* 107 (5):141–45.

———. 2019. "Interpreting signals in the labor market: evidence from medical referrals." *mimeo. Chicago* .

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2021. "Gender differences in recognition for group work." *Journal of Political Economy* 129 (1):000–000.

Schwarz, Carlo. 2022. "Estimating Text Regressions using textreg_train." *mimeo* .

Shepley, Carol Ferring. 2008. *Movers and Shakers, Scalawags and Suffragettes: Tales from Bellefontaine Cemetery*. Missouri History Museum.

Vaughn, Gerald 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.

Wigner, Eugene P. 1972. "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.

Appendix

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

- Appendix A provides details on the data collection.
- Appendix B shows additional results on hiring gaps.
- Appendix C shows additional results on publication gaps and a generalized version of the Roy model.
- Appendix D provides additional details on the predicted citations model.

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 country and rank. We recode them at the country level, because certain labels of ranks do not necessarily describe the same academic rank across countries. E.g., 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 jointly list a number of academics who hold different academic ranks (e.g., associate and assistant professors) without distinguishing the exact rank of each academic. In these cases, we assign the highest listed rank to each academic.

In many academic systems, e.g., in Germany and Italy, young researchers climb the academic ladder by substituting for full professors for some years and obtaining a professorship after that. For these countries, we code substitute professors as assistant professors.

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

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

2. associate professors (comprising the categories: associate professor, Emerita/us associate professors)

3. assistant professors (comprising the categories: assistant professors, Emerita/us assistant professor, and clinical faculty)

4. lower-ranked positions (comprising the categories: teaching position, research position).

Promotion to full professor. We classify academics who enter the data at ranks 2, 3, or 4 in cohort t and are promoted to rank 1 as promoted to full professor. Note, 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 different 100,000 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 Link Destatis for details).

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

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 within a city. We then decide whether two entries refer to the same academic based on the first name, specialization, academic rank, and title. In cases in which the information indicates that an academic holds two appointments, we harmonize, if necessary, 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.[44]

A.1.4 Linking Academics Across Cohorts

We link academics across cohorts t (1900, 1914, 1925, 1938, 1956, 1969, 2000). Linking academics over time is crucial to analyzing promotions.

The link allows for the possibility that academics report slightly different first names in two adjacent cohorts. Such variations in first names occur because of five main reasons:

1. Universities sometimes report first names with slight variations across cohorts. E.g., the University of Leipzig reported the geographer Joseph Partsch as Jose*ph* Partsch in the 1914 cohort but as Jose*f* 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. E.g., 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. E.g., the University of Chicago botanist *Henry Chandler* Cowles was reported as *Henry C.* Cowles in 1914 but as H. C. Cowles in 1925.

4. Some original names are Germanized or Englishized for some individuals in some cohorts. E.g., the Hungarian mathematician Gusztáv Rados was listed as *Gusztáv* Rados in 1925 but as *Gustav* Rados in 1938.

5. Name variations in the first name in rare cases may 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 In this part, we explain how we link academics who remain in the *same department* between cohort t and cohort $t+1$. In a first step,

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

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 the 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 entire information on the first name is identical in both cohorts, we consider these academics as linked (in some cases, the information on the first name that is reported for that particular academic may be one or more initials 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. E.g., the data contain the following data points for the cohorts 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 consider 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

In this part we explain how we link academics who remain in the *same country but change departments* between two cohorts. In a 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. 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.[45] We then process the potential links as follows:

[45] A small number of universities change the country over the time period we consider in our analysis. E.g., the University of Strasbourg is listed as a German university in 1900, and 1914, but as a French university

1. If the entire information on the first name is identical in both cohorts, we consider these academics as linked (in some cases the information on the first name that is reported for that particular academic may be one or more initials 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 valid, the research assistants use the following rules:

   (a) If there are minor spelling differences in the first name, the research assistant considers the potential link an actual link (see lines 1 and 2 in Table A.2)

   (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 an actual link (see lines 3 and 4 in 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 μ before the year corresponding to cohort $t+1$, the potential link is classified as valid

   E.g., K(arl) Röder (see lines 5 and 6 in Table A.2) could be found online (see Wikipedia Article) 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 Table A.2) they classify the potential link as an incorrect link.

from 1925 onward. Hence, the within country link for the University of Strasbourg considers academics who move from or to other German universities between 1900 and 1914 as well as 1914 and 1925. It also considers academics who move 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 changes the country) are considered in the cross-country link that we describe below.

Table A.2: **Examples: 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 In this part, we explain how we link academics who *move across countries*. In a 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. To rule out false positives, all potential links are confirmed by extensive manual online searches. Hence, 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. 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 an actual link.

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.[46] 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 the information on the same academic from a different cohort (see Appendix section A.1.4 on how we link academics across cohorts). E.g., the University of Chicago botanist *Henry Chandler* Cowles was reported as *Henry C.* Cowles in 1914 but as *H. C.* Cowles in 1925. We therefore adjust the first name in 1925 to *Henry C.* Second, we hand-check around 60,000 academics who are only reported with initials. For this step, research assistants google the initial(s), surname, discipline, and university to find online records for the respective academics.

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.

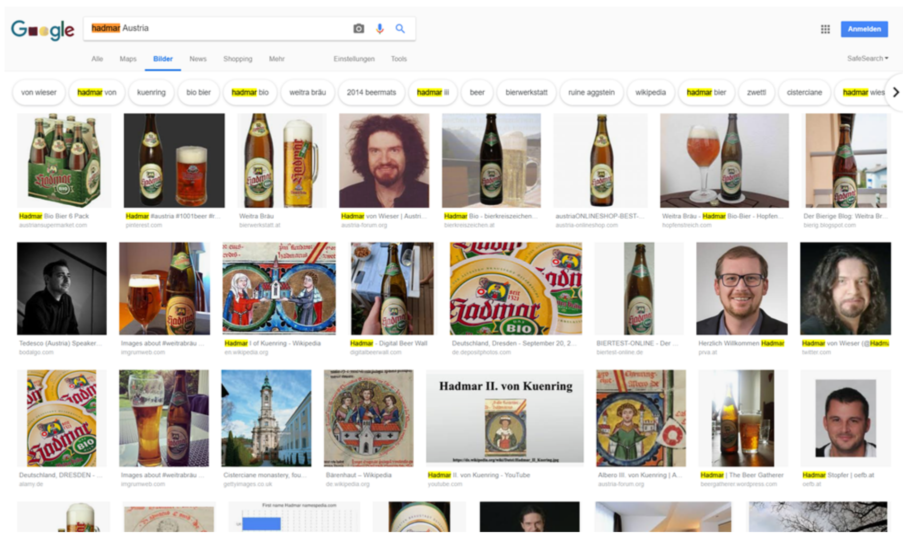
[46]For some academics, we can use the information on gender from the way that academics are listed in *Minerva* (e.g., as Miss or Mlle.) or from their website (e.g., from pictures or personal pronouns).

A.2 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 the first name by country combinations. Figure A.1 shows an example of the output of the *Google* image search, when we searched 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 name 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 combination (see section 1.1 for details).

A.2.2 Hand-Checking Gender Coding

As described in the main text, in the last step of the gender assignment, we hand-check individual academics who appear misclassified. Such misclassifications occur mostly because the predominant gender of a first name by country combination changes over time.[47] For example, French academics with the first name Camille were predominately 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 scientist and try to establish their gender. E.g., for the French biologist Camille Sauvageau, they find an entry in the *Proceedings of the Linnean Society of London* (from 1937) which says: "Camille Savageau

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

(1861-1936), Foreign Member of the Society, was born in Angers on 12 May 1861. *He* studied at Montpellier..." (see Link Google Books for details).

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

A.3.1 Homogenizing Author Names

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."[48] 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 by its 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." or "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).

3. We remove noble titles, e.g., "Della" or "Op Den" or "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 first name(s) initial(s) obtained in the previous step.

A.3.2 Preparing Addresses in *Web of Science*

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 cases, an alternative database, *Microsoft Academic Graph* (*MAG*), contains the relevant address information. We therefore enrich the

[48] In very rare cases, the *Web of Science* lists co-authors with identical surname and initial. Manual checks confirm that a lot of these are mistakes that occurred in the data entry by the *Web of Science*. We therefore drop one of the two observationally equivalent co-authors.

affiliations as reported by the *Web of Science* with information from *MAG*.[49] 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 affiliation "St. Louis, 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 affiliation, "St. Louis, USA", from the first paper to the second paper.

Processing Addresses with Google Maps

Over the very long time period that we study in this paper, some cities change their name (e.g., St. Petersburg became Leningrad) and a number of cities change 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 on a paper in the *Web of Science*. E.g. Rome is spelled using the German spelling "Rom" in *Minerva* or on the websites but spelled with either Italian ("Roma"), English ("Rome"), or German ("Rom") spelling in the *Web of Science*, depending on the country of the journal. To improve the match of papers to the faculty rosters, we therefore harmonize the address (in particular the country and city) in the *Web of Science* with the address in faculty rosters using *Google Maps*. The first step relies on the *Google Maps API*.

[49] *MAG* 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 around 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.

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*.[50] *Google Maps API* returns a JSON file that contains names of the city and the country, the centroid coordinates for the city, and a location-type flag that indicates the type of address that has been found (e.g. "CITY" if the *Google API* found a city). Similarly, we geocode the faculty roster data with *Google Maps API*. This also returns updated names of cities and countries. Crucially, as we process both addresses from the *Web of Science* and from the faculty rosters with *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 names 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*. These cases are easily identifiable as the location type flag is "APPROXIMATE" instead of "CITY". We improve the geocoding for these cases using the following 3-step procedure:

1. We structure the address before re-submitting it to the API, (e.g., "'city' : Preßburg, 'country' : Hungary").[51]

2. For those cases that did not return a result in step one, we provide a different structure (e.g., "Preßburg,+Hungary") of the address city and then re-submit it to the API.

3. For those cases that did not return a result in steps one and two, we re-submit the complete address 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 has been found. We therefore rely on the *Google Maps web interface*, as opposed to the *API*, for the 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 addresses that are misclassified, we calculate the Levenshtein distance between the city name in the *Web of Science* and the city name that *Google* returned. If the Levenshtein distance is larger than three (i.e., more than 3 letters differ), we copy the full address from the *Web of Science* into the *Google Maps* web interface. If the web interface

[50]The *Web of Science* already contains separate information on the city and country in addition to the full address.

[51]This option is not used as a baseline, since it reduces the match rate.

finds the address, we extract the city and country information 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 *Google Maps API* because the web interface returns somewhat different city and country names than the *API*.

The processing of addresses ensures that addresses from the faculty rosters and the *Web of Science* are harmonized and can then be matched as described in subsection A.4 below.

A.3.3 Predicting Academic Disciplines of Papers Using Paper Titles

To match papers from the *Web of Science* to the faculty rosters we also match on the discipline (subsection A.4 below). The *Web of Science* assigns papers to academic disciplines (e.g., physics, or general science) based on the journal they are published in, as opposed to assigning disciplines at the paper level. For 59% of the papers, this establishes a unique assignment to one of 8 disciplines (e.g., the journal *Acta Mathematica* is uniquely assigned to mathematics). The remaining 41% 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*).[52] Matching the latter papers to academics would involve considerable measurement error.

To uniquely assign disciplines at the paper level, independently of where the paper was published, we train a multinomial logistic regression classifier. This classifier, for example, assigns the more mathematical papers in *Biometrika* to mathematics while it assigns the papers with a biology focus to biology. We train the classifier is trained based on the words (unigrams), word pairs (bigrams), and word triplets (trigrams) from the titles of the 15,078,761 papers that the *Web of Science* already assigned to unique disciplines (e.g., the papers published in *Acta Mathematica*).

In preparation for the classifier, we remove very common words (stopwords) from the titles, as these do 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 the size of the vocabulary V is the total number of unique unigrams, bigrams, and trigrams in all titles.

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

[52]In the *Web of Science* (and in the faculty rosters) statistics is a sub-discipline of mathematics.

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 $tf{-}idf(w_{d,v}) = (1 + \log(w_{d,v}))\cdot\left(\log\left(\frac{1+D}{1+d_v}\right) + 1\right)$, 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 logistic regression 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, the optimal regularization strength is chosen using 10-fold cross-validation and evaluated on the basis of the F1-score.[53] 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. After the training process, we predict a unique discipline for the 10,508,299 papers which the Web of Science had originally assigned 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 data on scientists from the faculty rosters using a nine-step matching procedure. As mentioned in the main text, we match papers from the *Web of Science* within a ± five-year window around the year of the corresponding cohort of academics. E.g., for scientists listed in the 1914 cohort, we only match papers published between 1909 and 1919.[54] 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) subject, iv) country, v) city
2. Merge using: i) full surname, ii) all initials, iii) subject, iv) country, v) city
3. Merge using: i) full surname, ii) first initial, iii) subject, 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 false 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:

[53]The 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.

[54]See also footnote 20 in the main text.

(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 under step 3, we only consider matches where the full surname, subject, country, and city matches.)

(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. We then repeat steps 1-3, but remove the city as a merge criterion.

5. We repeat steps 1-3, but additionally remove the country as a merge criterion.

Note: 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 further 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 time period covered by our paper, the *Web of Science* rarely provides a unique assignment of the addresses reported on a paper to its co-authors: e.g., if a paper has two co-authors, and they are affiliated to different institutions, usually the *Web of Science* does not pin down which co-author is affiliated to which institution. We therefore merge each address reported in a paper to all of the co-authors of the paper. If there is more than one address associated with a paper, we perform a many to many merge of addresses to co-authors. As we show in Table 3 the results are robust to only considering scientists who have unique surname, first initial, and discipline in all universities of the world.

A.5 Benchmarking the Minerva Data

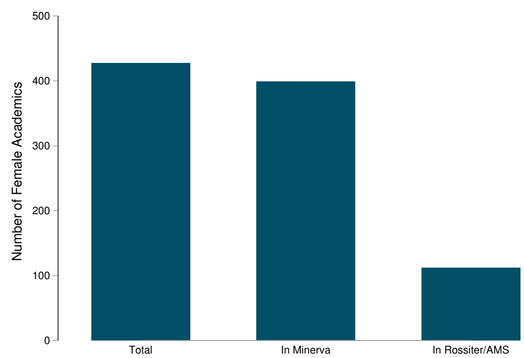
To the best of our knowledge, there are no comparable data that cover academics on a worldwide scale over the 20th century. Nonetheless, we can benchmark the *Minerva* data to smaller datasets that cover some universities and time periods.

A.5.1 Benchmarking Against Rossiter (1982) / American Men of Science (1938)

Rossiter (1982), pp. 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 that 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, Figure A.2). Of these, 399 (93%) are listed in *Minerva* (second bar).[55] 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*.

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.

A.5.2 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 in any German university during the winter semester 1937/38. The university calendar was published by J.A. Barth. He collected

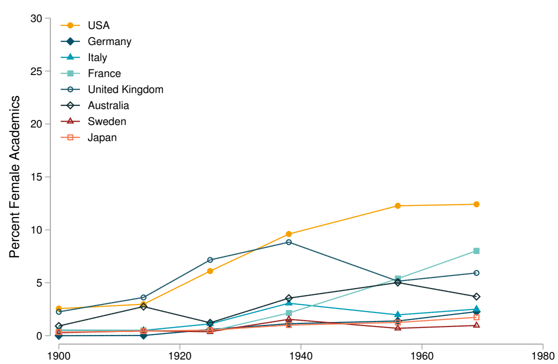
[55]The 7% missing female academics in *Minerva* are due to the following reasons: 1) in 1938 Minnesota (one of the 20 universities) only reported full professors in *Minerva* but Rossiter reports 9 female assistant or associate professors for Minnesota. 2) even though both sources were published in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

official university calendars from all 32 German universities and compiled them into one volume called *Kalender der reichsdeutschen Universitäten und Hochschulen*. We extract all physicists, chemists, and mathematicians in the same way as Waldinger (2012a). 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 scientists, a match rate of 98.5 percent, suggesting that the coverage of *Minerva* was very comprehensive.

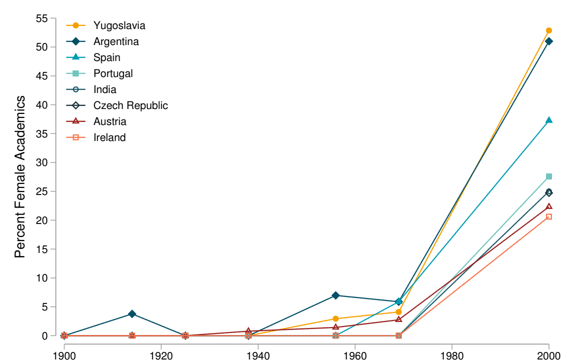
B Further Results: Hiring Gaps

Figure B.1: Percent of Female Academics by Country over Time, Additional Evidence

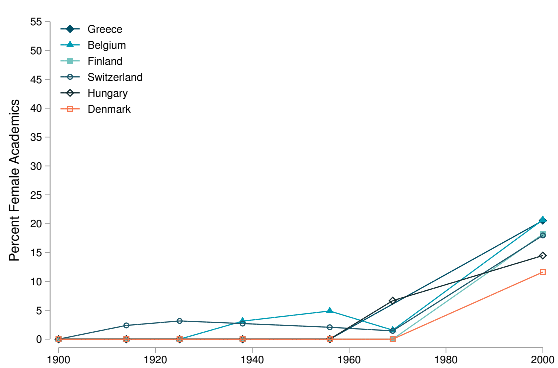
(a) Excluding Women's Colleges



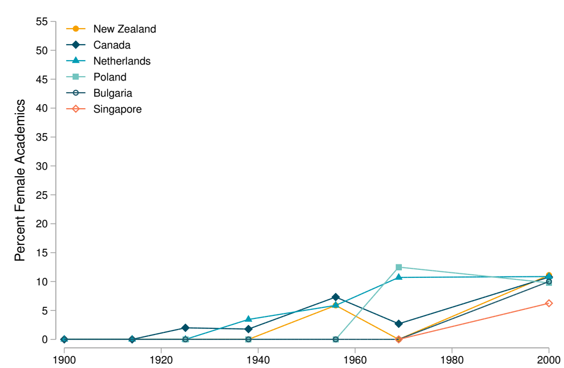
(b) Additional Countries



(c) Additional Countries



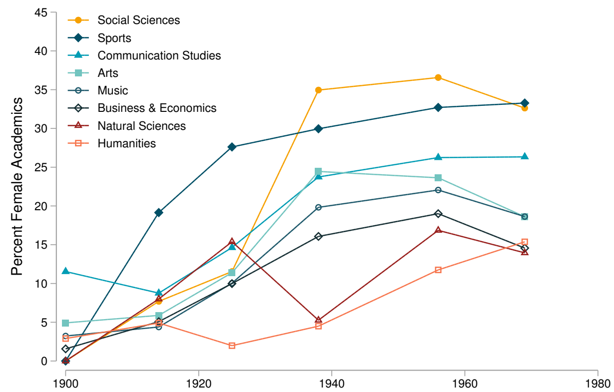
(d) Additional Countries



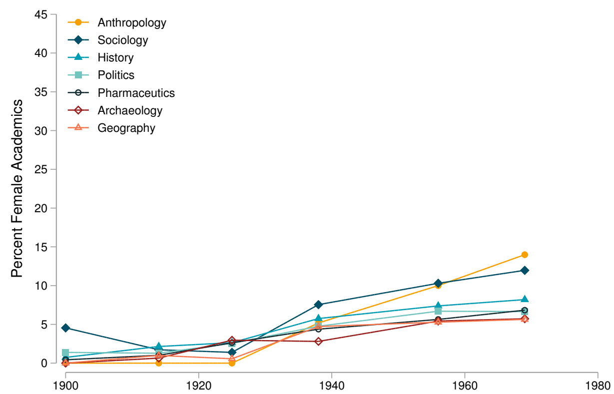
Notes: The Figure shows the percentage of female academics by country over time. Panel 1 plots the percentage of female academics across all universities and disciplines until 1969 (sample 1), excluding women's colleges. Panels 2-4 plot the percentage of female scientists in prestigious universities (sample 3). The data were collected by the authors from various volumes of *Minerva* and department websites, see section 1 for details.

Figure B.2: Percent of Female Academics, Additional Disciplines

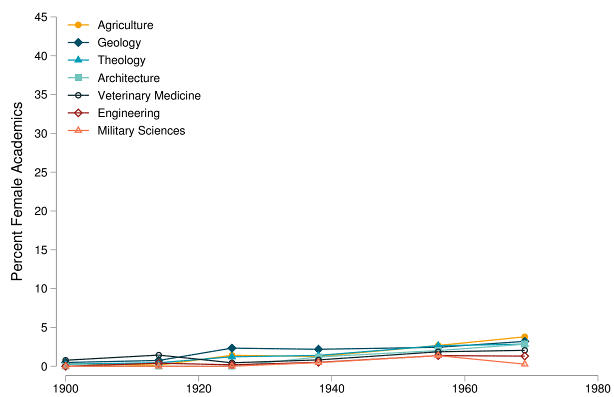
(a)



(b)



(c)



Notes: The Figure shows the percentage of female academics by discipline for additional disciplines not reported in the main paper for the period 1900-1969 (sample 1). Note that the figure combines Business and 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 Further Results: Publication Gaps

C.1 Additional Results: Publication Gaps

Table C.1: Individual-Level Publication Gaps (Additional Robustness)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	±Three-Year Window		Unique Matches		Full Professors		First Cohort	
Dependent Variable	Publi-cations	Standard. Publi-cations	Publi-cations	Standard. Publi-cations	Publi-cations	Standard. Publi-cations	Publi-cations	Standard. Publi-cations
Sample 2: All Universities 1900-1969								
Female (1900/14)	-1.055**	-0.210***	-1.610**	-0.222***	-0.447*	-0.170	-2.347	-0.213
	(0.482)	(0.055)	(0.753)	(0.065)	(0.241)	(0.136)	(1.871)	(0.137)
Female (1925/38)	-1.465***	-0.251***	-2.437***	-0.309***	-1.447***	-0.226***	-1.909***	-0.244***
	(0.505)	(0.043)	(0.751)	(0.050)	(0.510)	(0.072)	(0.589)	(0.043)
Female (1956/69)	-1.247***	-0.180***	-1.872***	-0.178***	-2.847***	-0.241***	-1.581***	-0.167***
	(0.408)	(0.026)	(0.652)	(0.033)	(0.984)	(0.061)	(0.535)	(0.030)
Observations	34,321	34,321	31,492	31,492	18,456	18,456	26,267	26,267
R-squared	0.351	0.232	0.357	0.257	0.437	0.326	0.416	0.306
Sample 3: Prestigious Universities 1900-2000								
Female (1900/14)	-1.349**	-0.227***	-2.089**	-0.247***	-0.314**	-0.089***	-3.116	-0.226
	(0.628)	(0.053)	(0.967)	(0.066)	(0.131)	(0.012)	(2.368)	(0.157)
Female (1925/38)	-2.186***	-0.321***	-3.819***	-0.433***	-2.294***	-0.296***	-2.919***	-0.302***
	(0.716)	(0.048)	(1.053)	(0.039)	(0.733)	(0.080)	(0.866)	(0.045)
Female (1956/69)	-2.791***	-0.379***	-4.003***	-0.372***	-6.489***	-0.506***	-3.776***	-0.405***
	(0.843)	(0.052)	(1.232)	(0.056)	(1.881)	(0.125)	(1.023)	(0.055)
Female (2000)	-2.076***	-0.190***	-3.162***	-0.202***	-0.213	-0.117**	-3.260***	-0.201***
	(0.520)	(0.032)	(0.796)	(0.033)	(1.568)	(0.057)	(0.824)	(0.035)
Observations	39,816	39,816	36,514	36,514	17,588	17,588	33,782	33,782
R-squared	0.276	0.127	0.280	0.144	0.387	0.212	0.286	0.148
Experience	Yes	Yes	Yes	Yes	Yes	Yes		
Cohort × Department FE	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 (mathematicians, chemists, and biochemists) in all universities covering the period 1900-1969. Sample 3 includes scientists in prestigious universities covering the period 1900-2000. In columns 1-2, a scientist's publications are counted in a ± three-year window around a cohort (i.e. 1911-1917 for a scientist listed in 1914). In columns 3-4, each scientist is defined as a unique lastname - first initial - discipline combination in every cohort and a scientist's publications are counted in a ± five-year window. In columns 5-6, only full professors are included in the analysis. In columns 7-8, only the first cohort in which a scientist is observed in the data is included in the analysis. In columns 1, 3, 5, and 7 the dependent variable equals the number of publications, while in columns 2, 4, 6, and 8 the standarized number of publications at the country-cohort-discipline level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman. The regressions control for cohort×department fixed effects. Standard errors are clustered at the discipline-country level. Significance levels: ∗∗∗ p<0.01, ∗∗ p<0.05, and ∗ p<0.1.

C.2 A Generalized Roy Model

In the more general version of the Roy model that we estimate, we relax assumptions (ii)-(iv) (see main text). In particular, the latent value of hiring a woman is specified as:

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

while that of hiring a man is:

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

where $r_g(\cdot)$ is an unknown function of observable characteristics X_i^g, $g \in \{W, M\}$, Δ_0 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:

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

with $\epsilon_{0i} = \epsilon_{0i}^W - \epsilon_{0i}^M$, $X_i = \left(X_i^W, X_i^M\right)$, and $r_0\left(X_i\right) = r_M\left(X_i^M\right) - r_W\left(X_i^W\right) + \Delta_0$. We assume that, while non-parametric with respect to X_i,[56] $r_0(\cdot)$ however satisfies the exclusion restriction that it does not depend on s_0^W, so that the gender bias in hiring is not a function of the share of women among all hired academics. The error term ϵ_{0i} is distributed according to F, an unknown c.d.f. with unbounded support and invertible. Given these, the share of female academics, conditional on observable characteristics X_i, can be expressed as:

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

Because observable characteristics X_i are unavailable on a world-wide scale over the 20th century, we assume $X_i = X_0$ for all i's, with X_0 some constant value such that $r_M\left(X_0^M\right) = r_W\left(X_0^W\right)$.[57] It then follows that equation (C.4) simplifies to $s_0^W = 1 - F\left(\Delta_0\right)$, representing the share of women among all hired academics.[58] This and the invertibility of F imply that $\Delta_0 = F^{-1}\left(1 - s_0^W\right)$, a fact we use below in equations (C.6) and (C.7).

Publication Market. In this more general version of the Roy model, we allow the gender bias in publications Δ_1 to be a function of the share of female academics, $\Delta_1\left(s_0^W\right)$. Conditional on academic position i being filled by either a woman or a man at the hiring stage, we observe the following outcome equations at the publication stage:

[56] For example, $r_0(\cdot)$ trivially fits the linear specification $r_0\left(X_i\right) = X_i\beta_0 + \Delta_0$ assumed in (5), but can also take the more general form $r_0\left(X_i\right) = g\left(X_i, \beta_0\right) + \Delta_0$, with $g\left(X_i, \beta_0\right)$ any function of X_i and the finite- or even infinite-dimensional parameter β_0.

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

[58] This clarifies the practical implication of the assumed exclusion restriction embedded in assumption (i) (see main text): if $r_0\left(X_i, s_0^W\right)$, even when $X_i = X_0$ for all i's, with X_0 some constant value such that $r_0\left(X_0, s_0^W\right) = r_0\left(s_0^W\right)$, (C.4) would still take the rather inconvenient form $s_0^W = 1 - F\left(r_0\left(s_0^W\right)\right)$.

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

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.

Publication conditional on gender. The expectation of Y_{1i}^W conditional on $Y_{0i} > 0$ from (C.5) is:

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

Analogously, the expectation of Y_{1i}^M conditional on $Y_{0i} \leq 0$ from (C.5) is:

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

Without further assumptions, it is impossible to separately identify the various components of $g_W(\cdot)$ and $G_M(\cdot)$.[59] To avoid unnecessarily strong functional form restrictions, we directly approximate these functions by polynomial expansions of the share of female academics: $g_W\left(s_0^W\right) = \sum_{\kappa=0}^{2} \theta_\kappa^W \times \left(s_0^W\right)^\kappa$ and $G_M\left(s_0^W\right) = \sum_{\kappa=0}^{2} \theta_\kappa^M \times \left(s_0^W\right)^\kappa$, respectively.[60] In practice, we estimate (C.5) and (C.6) on the basis of sample 3 (scientists, prestigious universities, 1900-2000) with a regression similar to (1) that also includes interactions with polynomials of the share of female academics:

[59] Assumptions (ii)-(iv) in the parametric version of this model overcome this lack of identification (see main text).

[60] Implementing these approximations with polynomials of higher degrees does not qualitatively affect our conclusions while leading to less precise estimates.

$$
\begin{aligned}
\text{Pub}_{idt} &= \gamma + \sum_{\kappa=0}^{2} \left(s_{0\ell(i)}^{W}\right)^{\kappa} \times \left(\gamma_{\kappa 1}\,\text{Female}_{idt} \times 1\left[t(i) = 1900/38\right]\right. \\
&+ \left.\gamma_{\kappa 2}\,\text{Female}_{idt} \times 1\left[t(i) = 1956/69\right] + \gamma_{\kappa 3}\,\text{Female}_{idt} \times 1[t(i) = 2000]\right) \\
&+ \text{Experience}_{idt}\boldsymbol{\gamma}_{\text{exp}} + \text{Fixed Effects} + \varepsilon_{idt},
\end{aligned}
\tag{C.8}
$$

where Pub_{idt} measures the standardized number of papers published by scientist i in cohort $t(i)$ and department d, $s_{0\ell(i)}^{W}$ is the share of female academics in scientist i's cohort-country $\ell(i)$ (e.g., USA in 2000), and each $\gamma_{\kappa p}$ corresponds to the portion of gender gap $\theta_{\kappa p}^{W} - \theta_{\kappa p}^{M}$, $\kappa = 0, 1, 2$, which we estimate separately for each period p. Figure 7 (see main text) summarizes the estimation results (in a regression with cohort $\times$ discipline $\times$ country and department fixed effects) by plotting, separately for each period p, the predicted gender gap in standardized publications $\widehat{g}_{W,p}\left(s_{0\ell}^{W}\right) - \widehat{G}_{M,p}\left(s_{0\ell}^{W}\right) = \sum_{\kappa=0}^{2} \widehat{\gamma}_{\kappa,p} \times \left(s_{0\ell}^{W}\right)^{\kappa}$ as a function of $s_{0\ell}^{W}$.

Table C.2: Individual-Level Publication Gaps and the Share of Females

Dependent Variable	(1) Standard. Publi-cations	(2) Standard. Publi-cations	(3) Standard. Publi-cations		(4) Standard. Publi-cations	(5) Standard. Publi-cations	(6) Standard. Publi-cations
	Inverse Mills Ratios, regression (11)				*Polynomial Approx., regression (C.8)*		
Female (1900/38)							
$-\Delta_{11}$	-1.085	-0.463	-1.090	γ_{01}	-0.358***	-0.233**	-0.241**
	(0.865)	(1.046)	(0.932)		(0.065)	(0.104)	(0.118)
ρ_1^W	0.250	0.055	0.281	γ_{11}	-2.337	-3.518	-4.450
	(0.302)	(0.379)	(0.334)		(2.472)	(2.809)	(3.120)
ρ_1^M	1.293	-0.027	1.069	γ_{21}	16.464	18.448	24.702
	(2.211)	(2.437)	(2.356)		(18.317)	(18.943)	(19.039)
Female (1956/69)							
$-\Delta_{12}$	-5.186***	-4.306**	-3.374	γ_{02}	-0.116	-0.127	-0.071
	(1.894)	(2.041)	(2.125)		(0.201)	(0.205)	(0.225)
ρ_2^W	1.952**	1.605*	1.256	γ_{12}	-9.562**	-8.406*	-8.199
	(0.789)	(0.855)	(0.889)		(4.459)	(4.591)	(4.965)
ρ_2^M	5.549**	4.452*	3.161	γ_{22}	31.187**	27.006*	25.143
	(2.486)	(2.617)	(2.711)		(14.902)	(15.308)	(16.366)
Female (2000)							
$-\Delta_{13}$	-2.526***	-1.978**	-1.745**	γ_{03}	0.182	0.200	0.171
	(0.885)	(0.792)	(0.753)		(0.169)	(0.142)	(0.128)
ρ_3^W	1.224***	0.980**	0.862**	γ_{13}	-3.207**	-2.974***	-2.715***
	(0.456)	(0.406)	(0.383)		(1.227)	(1.039)	(0.952)
ρ_3^M	1.505**	1.042*	0.861	γ_{23}	4.359**	3.774**	3.343**
	(0.685)	(0.625)	(0.608)		(1.856)	(1.604)	(1.506)
Observations	36,526	36,526	36,526		36,526	36,526	36,526
R-squared	0.020	0.064	0.123		0.020	0.064	0.123
Experience	Yes	Yes	Yes		Yes	Yes	Yes
Cohort × Discipline × Country FE	Yes	Yes			Yes	Yes	
Department FE		Yes				Yes	
Cohort × Department FE			Yes				Yes

Notes: The Table shows estimation results for Roy model (11) in columns 1-3, and for its more general version (C.8) in columns 4-6. Estimates are based on scientist-level observations from sample 3, which includes scientists in prestigious universities covering the period 1900-2000. The dependent variable equals the standardized number of publications in a ± five-year window around a cohort (i.e., 1995-2005 for a scientist listed in 2000), where the standardization is at the country-cohort-discipline level. In columns 1-3, the main explanatory variables are indicators that equal 1 if the scientist is a woman, and their interactions with inverse Mills ratios evaluated at $\pm\Phi^{-1}\left(s_{0\ell}^W\right)$, where $s_{0\ell}^W$ is the share of women in the cohort-country of the scientist. In columns 4-6, the inverse Mills ratios are replaced by second-degree polynomials of $s_{0\ell}^W$. All regressions exclude observations from cohort-country combinations that do not include any woman. The regressions also control for experience and various additional fixed effects, as indicated in the table. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

Table C.2 reports estimation results for both versions of the Roy model we consider, regression (11) in columns (1)-(3), and regression (C.8) in columns (4)-(6).

D Further Details on the Predicted Citations Model

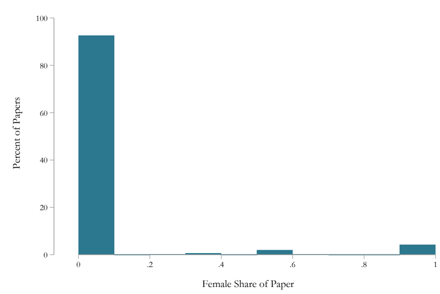
As outlined in the main text, we aim to account flexibly for the topic of each paper, which could influence the citations of the paper. We use a ridge regression model that uses the words (unigrams) and word pairs (bigrams) that appear in the title of the 364,282 scientific papers which we match to at least one scientist. The model learns how many citations, on average, papers in finely-grained research fields typically receive.[61]

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 $P \times V$, where D is the number of papers and V is the total number of unique unigrams and bigrams.

The model minimizes equation 12 to identify the n-grams that have the highest predictive power for citations. The regularization term λ reduces overfitting of the model to the training sample, by picking up individual n-grams that appear in some extremely successful papers. We choose the optimal normalization strength using 10-fold cross-validation. To incorporate differences in citations for papers published in different time periods and disciplines, we fit the model separately for each of our cohorts and discipline. The model can thus account for the changing importance of topics over time and across disciplines.

If women (a) wrote papers with different titles than men and (b) were systematically under-cited because of discrimination, the model could internalize discrimination against papers written by women if we trained it on all papers. To avoid such bias, we train the model on the 93% papers which were written by only male authors (see Figure D.1 for a histogram of the female share of the papers). We then use the estimated coefficients of equation 12 to predict citations for *all* papers (also those published by women).

Figure D.1: Histogram Female Share of Paper



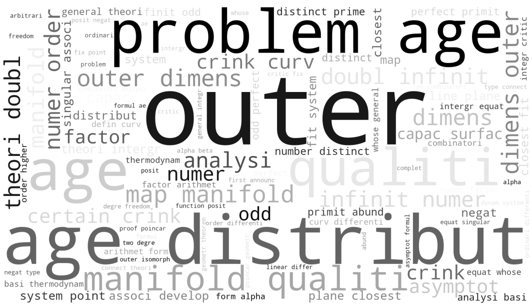
Notes: The Figure shows a histogram of the female share of papers for which we assign the gender for at least on of the authors.

[61] Hill and Stein (2022) use a similar approach based on information from the Protein Data Bank to train a machine learning model to predict citations of academic research.

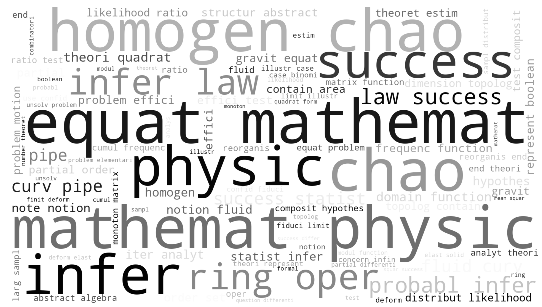
D.1 Additional Word Clouds Predicted Citations

Figure D.2: Words that Predict High Citations in Mathematics over Time

(a) Mathematics 1900-1914

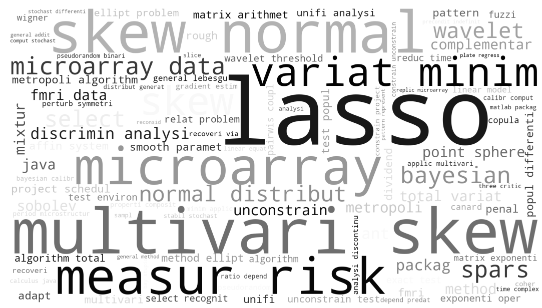


(b) Mathematics 1925-1938

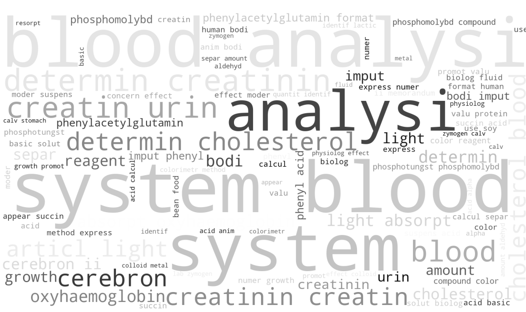


(c) Mathematics 1965-1969

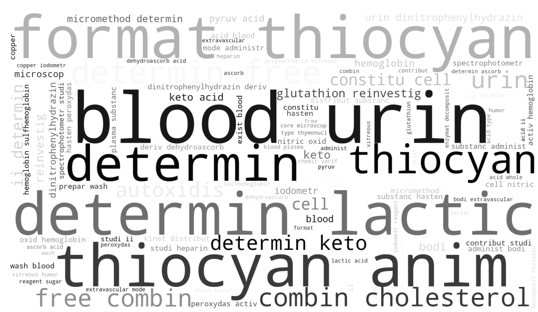
(d) Mathematics 2000



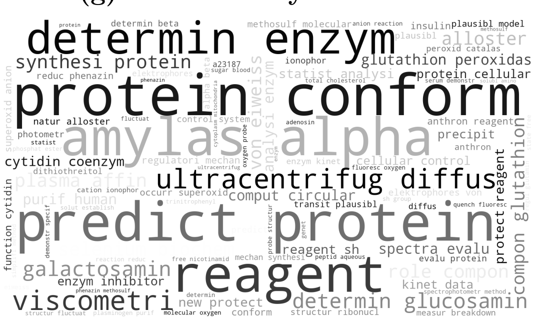
(e) Biochemistry 1900-1914



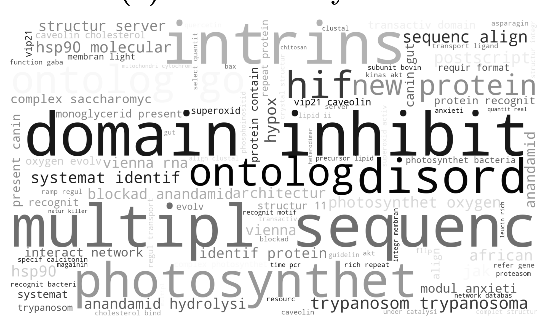
(f) Biochemistry 1925-1938



(g) Biochemistry 1965-1969



(h) Biochemistry 2000



Notes: The Figure shows the unigrams and bigrams that predict the highest citations in mathematics and biochemistry for the indicated cohorts. 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.

D.2 Additional Robustness Predicted Citations

Table D.1: Paper-Level Citations Gaps: Controlling for Predicted Citations - Robustness

	Baseline (1)	Linear Control (2)	Non-Parametric Control (3)	Out of Sample (4)	Without Winsorization (5)	Citation Count (6)
Sample 2: All Universities 1900-1969						
Female-authored Paper (1900/14)	-0.268**	-0.274**	-0.236**	-0.318***	-0.169**	-7.405***
	(0.106)	(0.106)	(0.096)	(0.093)	(0.073)	(1.608)
	[0.186]	[0.191]	[0.182]	[0.175]	[0.197]	[5.073]
Female-authored Paper (1925/38)	-0.141***	-0.141***	-0.130**	-0.163***	-0.073*	-6.493***
	(0.052)	(0.053)	(0.050)	(0.032)	(0.038)	(1.505)
	[0.074]	[0.076]	[0.079]	[0.056]	[0.055]	[3.126]
Female-authored Paper (1956/69)	-0.080***	-0.082***	-0.065***	-0.094***	-0.064***	-6.903**
	(0.025)	(0.025)	(0.023)	(0.026)	(0.022)	(2.901)
	[0.041]	[0.044]	[0.04]	[0.027]	[0.036]	[3.403]
Observations	155,264	155,264	155,264	78,939	155,264	155,264
R^2	0.443	0.435	0.511	0.057	0.466	0.776
Sample 3: Prestigious Universities 1900-2000						
Female-authored Paper (1900/14)	-0.341***	-0.351***	-0.317***	-0.365***	-0.197***	-9.084***
	(0.048)	(0.037)	(0.066)	(0.076)	(0.033)	(0.686)
	[0.13]	[0.138]	[0.139]	[0.137]	[0.137]	[7.025]
Female-authored Paper (1925/38)	-0.144***	-0.142***	-0.149***	-0.163***	-0.056	-6.416***
	(0.050)	(0.053)	(0.049)	(0.040)	(0.045)	(1.970)
	[0.078]	[0.079]	[0.084]	[0.066]	[0.061]	[4.188]
Female-authored Paper (1956/69)	-0.106***	-0.112***	-0.098***	-0.111***	-0.083***	-9.476*
	(0.031)	(0.033)	(0.029)	(0.032)	(0.028)	(5.410)
	[0.055]	[0.062]	[0.055]	[0.049]	[0.051]	[5.818]
Female-authored Paper (2000)	-0.016	-0.020	-0.011	-0.025*	-0.011	-0.109
	(0.014)	(0.013)	(0.014)	(0.014)	(0.013)	(0.894)
	[0.017]	[0.016]	[0.016]	[0.014]	[0.016]	[0.902]
Observations	274,447	274,447	274,447	148,339	274,447	274,447
R^2	0.377	0.360	0.406	0.085	0.374	0.703
Predicted Citations Control	Yes	Yes		Yes	Yes	Yes
Predicted Citations (1000 bins) FE			Yes			
Country × Discipline × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The Table shows gender gaps in citations. Results are estimated at the paper level. The dependent variable is the winsorized citation count which we standardize at the cohort-country-discipline level. The main explanatory variable and indicator variable which is equal to 1 if the paper has at least 1 female author. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for the first and second degree polynomial of the predicted citation variable or for 1000 dummies for the quantiles of the predicted citation distribution. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model. Standard errors are clustered at the discipline-country level. We additionally report boostrap standard errors in square brackets. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

References

Cattell, James McKeen. 1938. *American Men of Science: A Biographical Directory*, vol. 6. New York: Science Press.

Hill, Ryan and Carolyn Stein. 2022. "Race to the bottom: Competition and quality in science." *mimeo* .

Rossiter, Margaret W. 1982. *Women Scientists in America: Struggles and Strategies to 1940*, vol. 1. JHU Press.

Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June Hsu, and Kuansan Wang. 2015. "An overview of microsoft academic service (mas) and applications." In *Proceedings of the 24th international conference on world wide web*. 243–246.

Waldinger, F. 2012a. "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany." *Review of Economic Studies* 79 (2):838–861.