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

Alessandro Iaria (University of Bristol)

Carlo Schwarz (Bocconi University)

Fabian Waldinger (Ludwig-Maximilians-Universität München)

June 28, 2024

Abstract

Using the largest database of academics ever assembled, we analyze gender gaps in academia over an unprecedented time span and geographic coverage. First, we find that women were substantially less likely to be hired throughout the 20th century. Gender gaps in hiring differed across countries and disciplines, and declined over time. Second, women published fewer papers than men. Estimates of a Roy model show a U-shaped relationship between gender gaps in hiring and in publications, indicating that these gaps are inherently linked. With declining gender gaps in hiring, the relative importance of positive selection of women was offset by increased publishing opportunities for women. Third, women received fewer citations. We develop a novel machine learning approach that shows that citation gaps did not arise because women worked on less-cited topics. Fourth, women were less likely to be promoted to full professor, even accounting for gender gaps in publications and citations.

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

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, as of 2017, women occupied merely 5.8% of CEO positions in Fortune 500 corporations (Bertrand et al. 2019) and were granted only 9.5% of patents in OECD countries (OECD 2021). In academia, women remain under-represented in most countries. The sustained 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 academia. In particular, little is known about the long-run evolution of gender gaps across countries and disciplines. It also remains unclear in which domains (hiring, publishing, citations, promotions) gender gaps are more pronounced and if gender gaps interact across domains.

This paper makes progress in our understanding of gender gaps in academia by tracing their evolution across multiple domains over the 20th century and at a global scale. 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 more than half a million observations covering academics in 7,477 universities in all disciplines in 151 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 five 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. Among the many enhancements, 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. We also manually harmonize academic ranks across countries 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.

Because the data are based on complete faculty rosters, which we combine with data on publications and citations, we can study the entire population of academics who

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

[2]Margarete Bieber, an archaeologist and art historian, was the second woman in Germany to become a full professor. Because of her Jewish background, she was dismissed by the Nazi government and emigrated to the United States (Becker et al. 2023).

are, in principle, able to publish, be cited, and be promoted to full professor. This helps to overcome important selection biases that affect studies that exclusively rely on publication and citation databases.[3] This enables us to draw a more accurate portrait of the role of women in academia, which is not restricted to the sample of publishing academics. 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 gender 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, e.g., Katherine 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 227 women had been hired, a share of about 1%. In the following decades, the share of women increased slowly: 2% in 1914, 3% in 1925, 7% in 1938, 11% in 1956, and 11% in 1969. In the sciences (mathematics, physics, chemistry, biochemistry, and biology), female shares across all universities increased from about 1% to about 7% between 1900 and 1969. In prestigious universities, female shares in the sciences were about 25-50% lower (than in all universities) until 1969. Between 1969 and 2000, female shares in the sciences in prestigious universities increased substantially: from about 4.4% to 18.7%. Despite the increase over the last decades of the 20th century, women continue to be significantly underrepresented in the sciences at prestigious universities. Our data also indicate that throughout the 20th century, female participation in academia was consistently lower than in the broader workforce across most countries.

We further 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 114 women as full professors, a share of around 1%. In the following decades, the share of female full professors increased but remained always below the share among all academics. The slower increase in the share of women among full professors could either reflect compositional changes over time or worse career prospects for women. We investigate this question in the last part of the paper.

The global coverage of the data also enables us to uncover substantial heterogeneity in female shares 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

[3] A recent paper studying gender inequality in science based on publication data is Huang et al. (2020).

increased their shares substantially in the three decades until 2000. France and Italy had low shares before WWI, but had the highest female shares by 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, and possibly in other high-skilled professions, were affected by different factors from those in lower-skilled professions.

We also find substantial heterogeneity in female shares across disciplines. Averaged over the period 1900-1969, no discipline had a female share greater than 35%. Most disciplines had female shares below 10%. Gender gaps were substantially higher in STEM than in humanities or social sciences. Disciplines with particularly low female shares were physics, law, veterinary medicine, architecture, and theology. Disciplines with particularly high female shares were pedagogy, communication studies, languages, and sports sciences.

In the second part of the paper, we investigate gender gaps in output as measured by publications. One key advantage of studying academics is the availability of output measures that are comparable across time and space. Publications are one of the main 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. We measure publications over a ± 5 year interval around each cohort (e.g., 1995-2005 for scientists observed in 2000). In all universities, during the period 1900-1969, female scientists published on average one to two fewer papers than men (or around 0.2 standard deviations (s.d.)) over the ± 5 year interval. Gender gaps in publications are similar if we compare men and women in the same university and cohort, e.g., Harvard in 2000, or even in the same department and cohort, e.g., physics in Harvard in 2000. In prestigious universities, we estimate around 50% larger gender gaps in publications. We also estimate the evolution of gender gaps in publications over the 20th century. The publication gap was about 0.2 s.d. in 1900. In the following decades, the gender gap in publications increased and reached a maximum of 0.45 s.d. by 1956. After that, the gap declined to about 0.2 s.d. by 2000.

This over-time pattern of gender gaps in publications and the findings in the first part of the paper, which show an increase in the share of women over the 20th century, suggest that changes in the representation of women in academia may be related to relative changes in the selection and publishing opportunities of men and women. We propose a model along the lines of Roy (1951) to study whether changing gender gaps in hiring affect gender gaps in publishing. 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 make a scientist's publication output a function of the share of women in the profession 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.

Estimates of this model reveal a U-shaped relationship between gender gaps in publications and the share of female scientists. The U-shaped relationship appears across different time periods and scientific disciplines and is robust to estimating the model with different sets of fixed effects, which control for persistent differences across cohorts, countries, disciplines, and even universities or departments. The estimated U-shaped pattern (which we call, in short, gender U) suggests that gender gaps in hiring and gender gaps in publishing are inherently linked: as the share of women in science increased, the relative selection, as well as publishing opportunities, of men and women may have evolved systematically. A plausible interpretation for the downward-sloping part of the gender U is that the most talented women entered academia first (selection effect). With very few women in the profession, only Marie Curie and similarly brilliant women were hired (the "Marie Curie periods").[4] With increasing shares of women in the profession, gender gaps in publications became more negative. The upward-sloping part of the gender U may arise because higher shares of women in academia were accompanied by improved publishing opportunities for women (empowerment effect).

In the third part of the paper, we explore whether papers published by women received fewer citations. We propose a novel machine learning approach to investigate whether potential citation gaps stemmed from gender differences in research topics. We estimate a regularized regression using the words in paper titles to predict the expected number of citations for each paper. We use two approaches to predict citations. For the first approach, the training sample consists of all papers in our data. For the second approach, the training sample solely consists of papers published by men, predicting the actual citations of each paper as if it had been published by men. A model trained on all papers may give a better prediction of the actual citations. In contrast, a model solely trained on papers by men would alleviate the concern that, for any given paper title, citations of women may be downward biased because of discrimination.

Before WWI, papers by female authors received around 0.2 s.d. fewer citations than those by male authors. In the interwar and post-war periods, this gap reduced to around 0.1 to 0.15 s.d. By 2000, the gender gap in citations declined to about 0.05 s.d. The estimated gender gaps in citations are very similar if we control for the predicted citations and other characteristics of the paper. This indicates that gender gaps in citations do not stem from gender differences in the number of coauthors, the journals in which women publish, and most importantly, the topics that women work on. Instead, papers by women received fewer citations because of biases in citing behavior.

In the fourth part of the paper, we investigate how women advanced in their academic careers by studying gender gaps in promotions. We find that women were around 10-20 percentage points less likely to be promoted to full professor until the late 1960s.

[4]Rossiter (1982), p. 130, 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]."

By the year 2000, the gender gap in promotions in the sciences persisted but had closed to about 7 percentage points. Notably, gender gaps in promotions are very similar if we compare men and women who entered the data in the same university and cohort (e.g., Berkeley in 1900) or even the same department and cohort. This suggests that gender gaps in promotions throughout the 20th century were not attributable to women starting their careers in universities or departments with worse career prospects.

A potential explanation for women's lower promotion prospects could be the gender gaps in publishing and citations that we uncover in parts two and three of the paper. When we control for each scientist's publication and citation record, we estimate very similar gender gaps in promotions. This suggests that the gender gap in promotions was driven by other biases. Strikingly, this unexplained gender gap in promotions is larger than the effect of a three s.d. worse publication record.

In summary, we show the existence of substantial gender gaps in hiring, publications, citations, and promotions over the 20th century. We find that all four gender gaps have declined over the course of the 20th century, especially between 1969 and 2000. However, gaps remain substantial in all domains. Our analysis also indicates substantial heterogeneity in gender gaps across countries and disciplines. We also show that gender gaps in different domains are interconnected and uncover a U-shaped relationship between the gender gaps in publishing and hiring. The documented barriers that excluded women from participating may have resulted in "lost Marie Curies," depriving the academic community of valuable ideas and potential breakthroughs.[5] In a world where ideas play an ever-increasing role, this may have slowed down scientific progress and ultimately economic growth (e.g., Romer 1986; Romer 1990; Jones 1995).

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. 2022); women were less likely to be nominated as Fellows of the Econometric Society, the National Academy of Science, and of the AAAS until the late 1970s, but more likely to be nominated since the mid-2000s (Card et al. 2023); women receive less credit for group work (Sarsons 2017b; 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). Investigating subjects other than 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 2021); that women are more likely to perform tasks with low promotability (Babcock et al. 2017); that a higher share of women in evaluation committees lowers promotion prospects of women (Bagues et al. 2017); that physicians become more pessimistic about female sur-

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

geons' ability after a patient's death (Sarsons 2017a); and that a higher share of female undergraduates shifts the research topics of professors (Truffa and Wong 2022). 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 more than 100 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 et al. 2010), executives (e.g., Bertrand and Hallock 2001; Gayle et al. 2012; Albanesi et al. 2015), lawyers (Azmat and Ferrer 2017), pharmacists (Goldin and Katz 2016), and engineers (Roussille 2021); all in the United States. Our new database enables us to trace the evolution of gender gaps for one high-skilled profession at a global scale 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.[6]

A nuanced understanding of gender gaps sheds light on the many failures and the few success 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" (Minerva) and modern university websites. We combine these data with detailed publication and citation records from Clarivate Web of Science and Microsoft Academic Graph. Throughout the paper, we present results for three samples:

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

[6]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 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; Goldin 1990.)

Sample 1 allows us to study gender gaps in hiring and promotions in all disciplines and universities worldwide. Sample 2 which focuses on the sciences, additionally allows us to study gender gaps in publications and citations because, already by 1900, the sciences had developed a culture of publishing and citing research that was consistent across countries and similar to today's standards. Lastly, sample 3 extends the analysis until 2000 for the prestigious universities. Access for women to these prestigious institutions is of particular importance because a large share of scientific discoveries are made in these institutions.

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. 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.[7] 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).[8] For the remainder of the paper, we refer to these years as cohorts. For the digitization, we scan all pages of the relevant volumes and process them using optical character recognition (OCR) software. 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 Rome. Virtually all Ph.D. granting institutions are included in the data. For example, the data contain academics in 1,540 universities in the United States, 309 universities in the United Kingdom, 281 in Germany, and 351 in France (Appendix Table A.3 reports the number of universities for all countries).

[7] An article in Nature compared Minerva to the French Annuaire Général des Universités and noted that "[i]n scope...this annual is akin to the well-known 'Minerva...'. It is, however, very much less exhaustive" (Nature 1930). To the best of our knowledge, there are no comparable data covering academics on a worldwide scale over many decades. We benchmark Minerva in two ways. First, we show that the number of universities that are covered in Minerva is similar to the universities included in the World Higher Education Database (WHED (2024), see Appendix Table A.3). As the WHED does not include microdata on individual academics, we perform additional benchmarking exercises on smaller datasets that cover individual academics in some universities and time periods. The benchmarking exercises suggest that Minerva indeed covered a large fraction of the world's academics (Appendix A.6.).

[8] 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 cohorts were chosen based on data availability and to follow important historical events that affected universities during the 20th century: 1900: start of Web of Science data, 1914: cohort before WWI, 1925: cohort at the end of the boycott against academics from Central powers (Iaria et al., 2018), 1938: cohort before WWII, 1956: first post-WWII cohort compiled by Minerva, 1969: last cohort compiled by Minerva, 2000: first year with wide-spread coverage of academics on department websites.

Minerva lists the name of the university, followed by faculty rosters. For most universities, the data list assistant, associate, and full professors, but also honorary professors, and in some cases research positions and teaching positions.[9] 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 7,477 universities in more than 130 countries. Appendix Figure A.2 shows the global distribution of academics across cities.[10]

Modern Faculty Rosters For The Year 2000

For the year 2000, we digitize faculty rosters from archived university websites from the Internet Archive Wayback Machine. We focus on five disciplines: mathematics, physics, chemistry, biochemistry, and biology, and collect faculty rosters for 249 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.[11] Between 1900 and 1969, approximately 51% of all scientists who published 68% of papers in the five disciplines were affiliated with these 249 institutions. Thus, these prestigious universities represent a significant portion of worldwide scientific activity.

Enhancements of Faculty Roster Data

We make a large number of manual enhancements to the faculty rosters (see Appendix A.1.). First, we manually recode thousands of 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 specializations (e.g., "quantum theory" or "physique des particules élémentaires") into 36 disciplines (e.g., physics; see Appendix A.1.2.). Third, if academics hold multiple positions in 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 cohorts using a cascading procedure (see Appendix A.1.4.). Fifth, for academics listed only 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.).[12] Sixth, we construct consistent university identifiers by linking universities across cohorts and tracking mergers and splits over the 20th century.

[9] 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 498,525 faculty members with names (Table 1) and 108,398 additional faculty members (e.g., the 10 lecturers) without names.

[10] 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. (2023) contains 47,897 academics in 198 universities covering the period 1000 to 1800.

[11] The sample contains 69 universities in the United States, 31 in Germany, 24 in the United Kingdom, 22 in Italy, 21 in France, 9 in Switzerland, 7 in Australia, Austria, and Canada, 5 in Belgium and the Netherlands, 4 in Denmark and Sweden, 3 in Japan and Ireland, 2 in Argentina, Finland, Hungary, New Zealand, Portugal, and Spain, and 1 in Bulgaria, Chile, Croatia, Czech Republic, Greece, India, Israel, Norway, Pakistan, Peru, Poland, Romania, Russia, Serbia, Singapore, and Uruguay.

[12] All results remain unchanged without this step.

Table 1: Summary Statistics

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

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

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 (pictures and 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 pairs with gender-api.com,[13] which assigns a gender probability to 'first name'-country pairs.

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

In the fourth step, we process the remaining cases that gender-api.com classified as less than 100% male by searching the 'first name'-country pairs using a Google image search. A research assistant then classifies each 'first name'-country pair as male or female depending on whether the image search returns more male or female individuals.

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

E.g., gender-api and the research assistants could not identify the gender of "Hadmar" in Austria. We thus performed a Google search for "Hadmar Austria" and analyzed the resulting images. In this example, the images depicting individuals showed only men (see Appendix Figure A.1). We, therefore, code Austrian scientists called "Hadmar" as male.

In the fifth step, we hand-check individual academics who appear misclassified with a Google search (see Appendix A.2.2.). Misclassifications mostly occur because the predominant gender of some names changed over time. E.g., the French name "Camille" can be both male and female. In the early cohorts, most academics called "Camille" are males, while in later cohorts some are females.[14] While the manual steps significantly increase data quality, none of the results change without steps 3-5. All results in the paper use the sample of academics for whom we determine gender (see Table 1). Overall, we assign gender to 83.9 percent of the academics. The vast majority of the academics that cannot be assigned to a gender are in universities that only report academics with initials.

Examples of Academics in the Database

Figure 1 shows three exemplary academics for each cohort. The selection showcases some of the data's country, discipline, cohort, and gender dimensions. However, it does not do justice to the tens of thousands of academics who have contributed to the progress of knowledge. For 1900, the data include the economist Alfred Marshall (University of Cambridge), the physicist and Nobel laureate Max Planck (University of Berlin), and the sociologist Max Weber (University of Heidelberg).

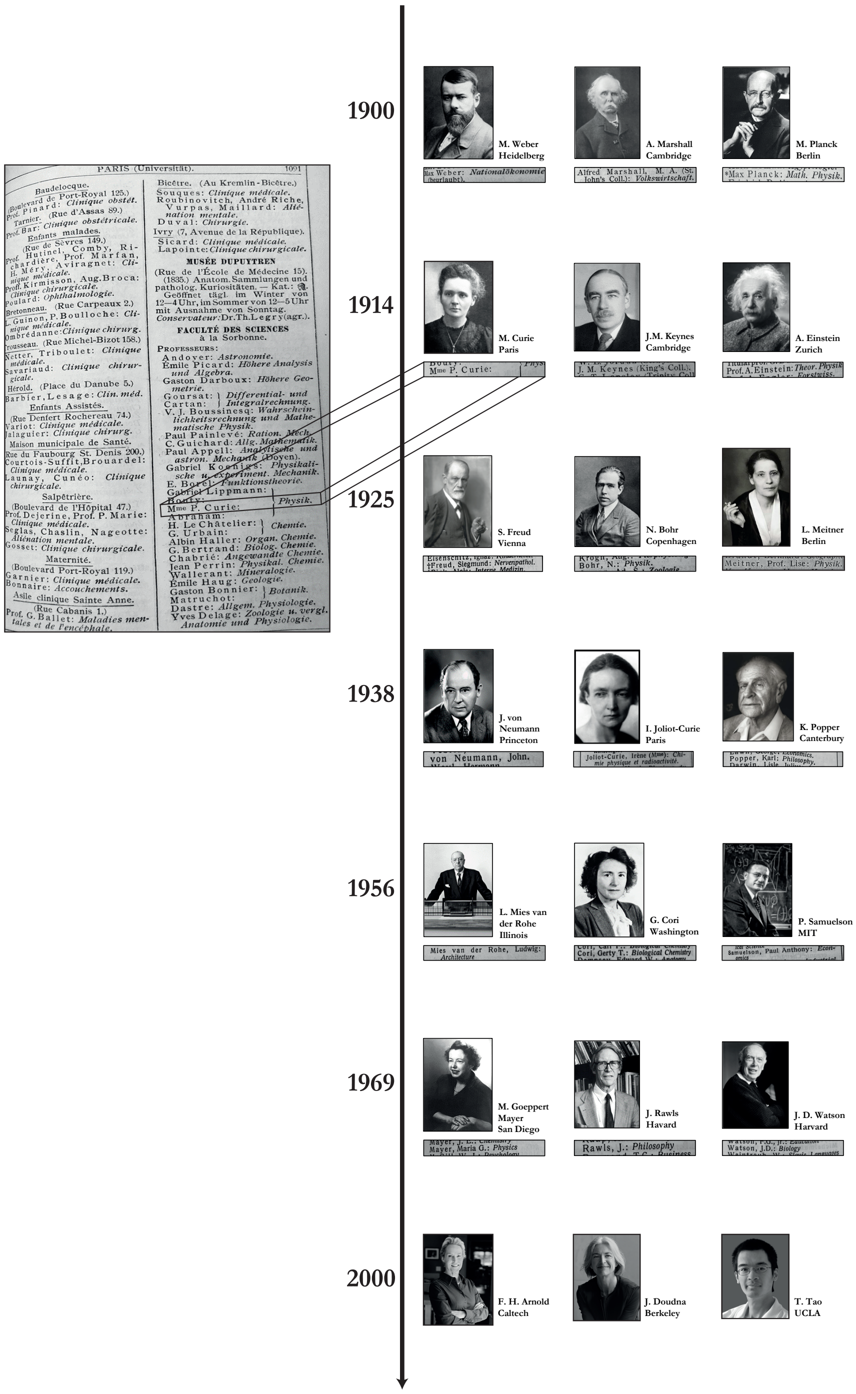
Examples for 1914 are John Maynard Keynes (University of Cambridge), Albert Einstein (ETH Zürich), and arguably the most famous woman in our data, Marie Curie (Université de Paris). Together with her husband, she conducted pioneering research on radioactivity and was the first woman to win the physics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

In 1925, the data list the physics Nobel laureate Niels Bohr (University of Copenhagen), the founder of psychoanalysis Siegmund Freud (University of Vienna), and Lise Meitner (University of Berlin). Meitner was the second woman to earn a physics PhD at the University of Vienna. During her Post-Doc at the University of Berlin, she was unpaid and had to run her experiments in a converted carpenter's shop in the cellar because — as a woman — she was barred from entering the main laboratory. Together with Otto Hahn she discovered nuclear fission. The Nobel laureate Wolfgang Pauli commented

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

[15] Strikingly, Curie is listed in Minerva as Mme P[ierre] Curie at a time when she had won two Nobel Prizes (Figure 1). Throughout her career, she faced obstacles because of her gender. In 1911, her nomination to the Academy of Sciences was met with resistance, epitomized by physicist Émile Amagat's argument "Women cannot be part of the Institute of France." Despite the efforts of some of France's greatest scientists, she narrowly lost the membership election to a male competitor (Curie 1938, pp. 277).

Figure 1: Examples of Academics in Database



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

that "Hahn and Meitner were great friends, but when they talked, she was superior." In 1945, the Nobel Prize was awarded to Hahn alone, neglecting Meitner's role, a decision contemporaries called a "stupidity of the Swedish Academy." (Kricheldorf 2014, p. 219).

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

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

The 1969 cohort includes the 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 chemistry Nobel laureate Frances H. Arnold (Caltech) and the biochemist Jennifer Doudna (Yale), who co-discovered a method of gene editing using CRISPR/Cas9. Together with Emmanuelle Charpentier, she was awarded the Nobel Prize in 2020 — the first all-female winners of the chemistry Nobel Prize. Another example from the 2000 cohort is the mathematician Terence Tao (UCLA), who won the Fields Medal in 2006.

1.2 Publication and Citation Data

To study gender gaps in publications and citations, we augment the faculty rosters with publication and citation data from Clarivate Web of Science. For any result based on publications and citations, we focus on five academic disciplines: mathematics, physics, chemistry, biochemistry, and biology. There are three reasons for this. First, these disciplines have particularly good coverage in the Web of Science. Second, they had already established the culture of publishing in scientific journals by 1900, and the publishing process was similar to today's. Third, the publishing process was international (Iaria et al. 2018). For the years of our study, the Web of Science contains papers in 14,191 journals in these disciplines. Naturally, the coverage of the Web of Science is not uniform across countries, disciplines, and over time. This does not affect our estimates as we control for cohort-discipline-country (or finer) 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.[16] To har-

[16]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.

monize affiliations across the faculty rosters and the Web of Science, we rely on Google Maps API (see Appendix A.3.2.) to extract cities and countries for each of the hundreds of thousands of unstructured affiliations. E.g., we extract the city "Cambridge" and the country "UK" from the affiliation "Cavendish Lab., Cambridge University, UK."

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

We consider publications in a ± five-year window around the year of the corresponding cohort. E.g., for scientists in the 2000 cohort, we match papers published between 1995 and 2005.[17] 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.[18]

2 Gender Gaps in Hiring

Hiring of Women Over Time

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

Across all countries and disciplines (sample 1), our newly collected data show that in 1900 only 227 women had been hired, a share of about 1% (Figure 2, panel a). In the following decades, the share of women increased, in particular between 1925 and 1938, i.e., before WWII. By 1969, a total of 17,204 women worked across all universities and disciplines, a share of about 11% — still nowhere close to equal representation.

We also explore changes in female shares among full professors. All over the world, full professor is the highest academic rank, which guarantees unique privileges and par-

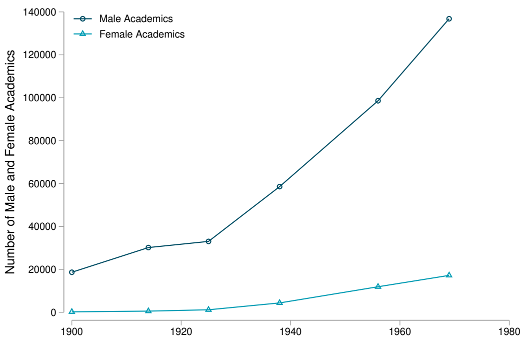
In some of these cases, Microsoft Academic Graph (MAG) contains the relevant affiliation. We, therefore, enrich the affiliation information with data from MAG (see Appendix A.3.2.).

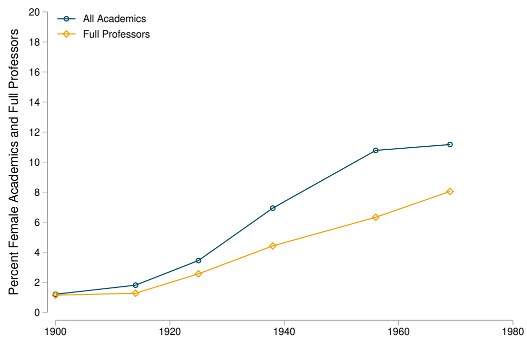
[17] We use a ± five-year window because scientists do not necessarily publish every year. Concerns that the matches of female academics may be affected by surname changes from marriage are mitigated by various factors. First, the faculty rosters list academics who are at least assistant professors.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

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

Figure 2: Absolute and Relative Number of Female Academics over Time

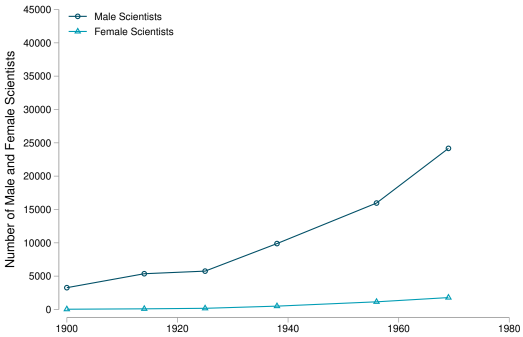
(a) *Sample 1:* All universities, all disciplines, 1900-1969

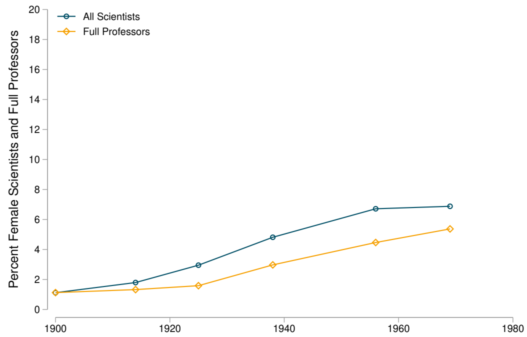




All Academics **Share female: All and Full Professors**

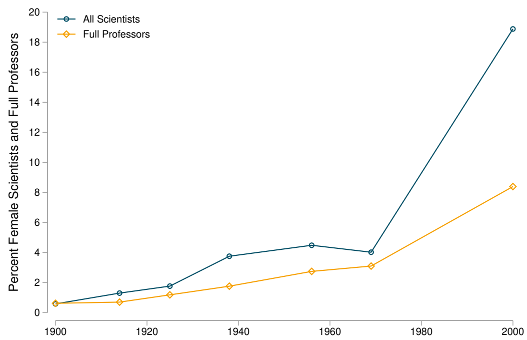
(b) *Sample 2:* All universities, sciences, 1900-1969





All Academics **Share female: All and Full Professors**

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



All Academics **Share female: All and Full Professors**

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

ticularly high job security and salaries. In addition, full professor is the most comparable academic rank across different university systems. In 1900, only 114 women worked as full professors across all universities, representing about 1%. In the following decades, the share of women among full professors increased, and by 1969 reached about 8%.

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

In the sciences (mathematics, physics, chemistry, biochemistry, and biology), female shares were substantially lower than for all disciplines. Across all universities, the share increased from about 1.1% to about 6.6% between 1900 and 1969 (Figure 2, panel b). In prestigious universities, female shares in the sciences were even lower until 1969. This suggests that women faced particularly large hurdles to obtain positions in prestigious universities. Between 1969 and 2000, female shares in the sciences in prestigious universities increased substantially: from about 4.4% to about 18.7% and from 3.3% to 8.8% among full professors (Figure 2, panel c). Despite this large increase, women were still heavily underrepresented in prestigious universities in 2000, especially among full professors.

In general, these patterns show substantial gender gaps in hiring throughout the 20th century. These gaps arose because women faced barriers throughout their upbringing and their educational careers. Parents may have treated young boys differently from young girls, and young girls may have been sent to schools with less academic-oriented curricula. In the first half of the 20th century, women faced substantial barriers to enrolling in universities, and even more so in PhD programs (Rossiter, 1982). Barriers may also stem from gender differences in exposure to academic role models (Bell et al., 2019). Finally, women may have faced discrimination when applying for faculty positions. Disentangling the importance of each of these channels is important to address hiring gaps. However, data on each of these channels covering 151 countries over the entire 20th century are currently unavailable. This makes the task of disentangling these channels beyond the scope of any one paper, but will hopefully be undertaken in future research.

Hiring Gaps Across Countries

The aggregate statistics hide significant heterogeneity between geographical regions. Until 1925, North America had much higher female shares than other regions. By 1938, the female share in the few African universities had caught up. Asia and South America had very low female shares until WWII, but then increased to the level of North America and Africa. Europe and Oceania had very low female shares until 1969. The variation in female shares at the country level was even higher, which we document by showing a selected set of countries that employed the largest number of academics throughout the 20th century. Before WWI, universities in the United States and, to a lesser extent, the United Kingdom, hired more women than any other country in the world. In the sample of all universities and disciplines (sample 1), the dominant role of the United States persisted until 1969 (Figure 3, panel a).[19] In the sciences (sample 2), universities in the United States and the

[19]The early U.S. lead is partly explained by women's colleges hiring more women. However, by 1938, other U.S. universities were also hiring a higher share of women than universities in other countries

United Kingdom hired a higher share of women than other countries until WWII. They were then overtaken by France (Figure 3, panel b). In prestigious universities (sample 3), only US universities had hired any woman by 1900 while in the other seven countries no prestigious university had hired a single woman. By 1914, prestigious universities in the United Kingdom had hired a similar share of women as the United States. The early lead of the United States and the United Kingdom lasted until 1938, when other countries started overtaking the Anglo-Saxon countries (Figure 3, panel c).

Sweden, and to a lesser extent Germany, had very low female shares until 1969 but increased their shares substantially in the three decades until 2000.[20] Italy and France had low shares before WWI, were ranked in the middle until 1969, but then increased their shares substantially in the three decades until 2000.[21] Japan is a clear outlier: female shares in prestigious universities were very low during the first decades of the 20th century but, unlike the other advanced countries, female share only increased slightly until 2000.

There are no comparable data that document gender gaps in hiring at the global level for the entire 20th century, neither for specific occupations nor for the entire labor force. However, we correlate gender gaps in academic hiring to female labor force participation for selected countries and periods (Appendix Figure B.2). We find no systematic relationship with trends in female employment in the general population. This suggests that the careers of women in academia evolved differently from 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 (Shanghai Ranking 2020). We choose the highest-ranked universities in each country and report the average female shares from 1900 to 2000.[22]

We report data on universities from various countries: ten universities from the United States; five each from Germany and the United Kingdom, three each from Canada, Japan, and Switzerland; two from France and Italy; and one from Argentina, Australia, Austria, Belgium, Denmark, Finland, Ireland, the Netherlands, Norway, and Sweden. The figure shows large differences in female shares across universities. Even within countries, university-level female shares vary widely. E.g., over the 20th century, Columbia hired, on average, around 8% of women in the sciences, while Princeton only hired around 2%. (Appendix Figure B.1).

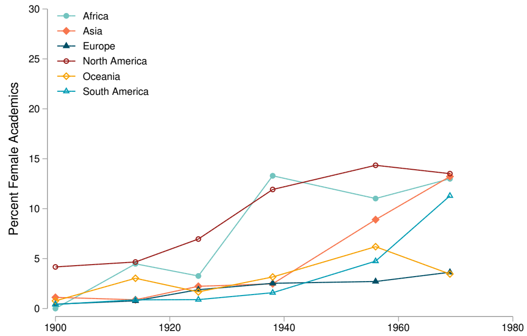
[20] Austria and Finland show a similar development, as shown in Appendix Figure B.1.

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

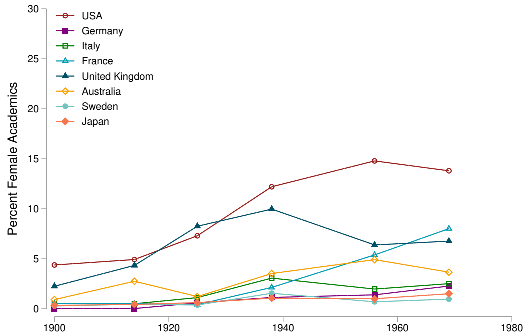
[22] The Shanghai Ranking ranks universities as of 2020. In many countries, e.g., the United States, the ranking has remained stable since 1900. In other countries, the ranking has changed substantially. To accurately reflect the most important institutions during the 20th century, we deviate from the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

Figure 3: Percent of Female Academics by Country over Time

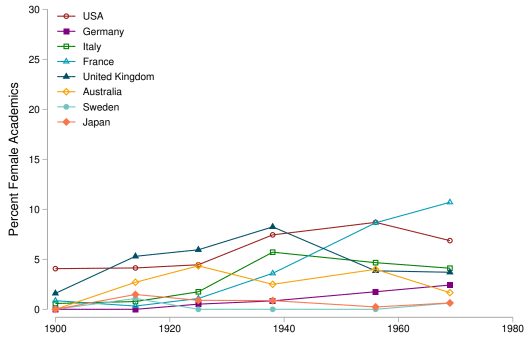
(a) *Sample 1:* All unis, all disciplines, 1900-1969



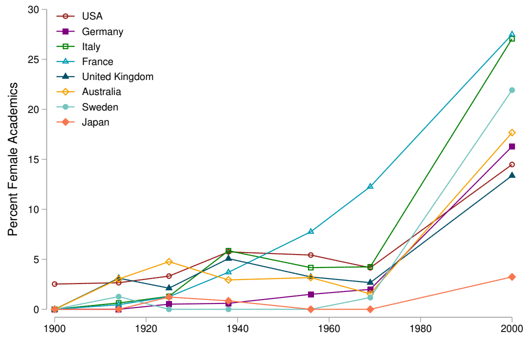
(b) *Sample 1:* All unis, all disciplines, 1900-1969



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

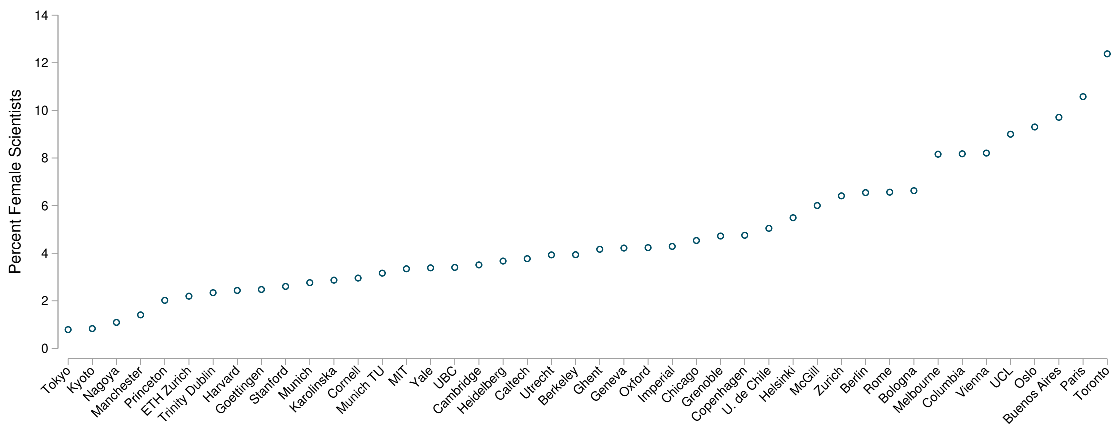


(d) *Sample 3:* Prestigious unis, sciences, 1900-2000



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

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



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

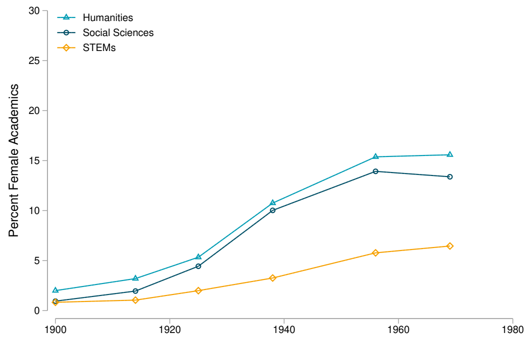
Hiring Gaps Across Disciplines

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

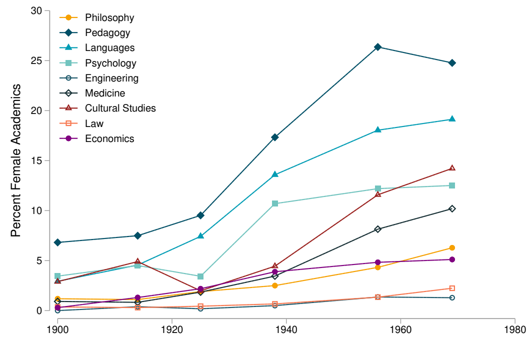
At a disaggregate level of disciplines, female shares varied substantially. In the first decades of the 20th century, most disciplines had very low (below 5%) female shares. For most disciplines, female shares remained below 10% until 1969, with particularly low shares in law, physics, and philosophy. However, some disciplines had higher female shares, which increased to about 25% by 1969 in pedagogy and 20% in languages (Figure 5, panel b). Appendix Figure B.3 shows additional disciplines.

Figure 5: Percent of Female Academics by Discipline over Time

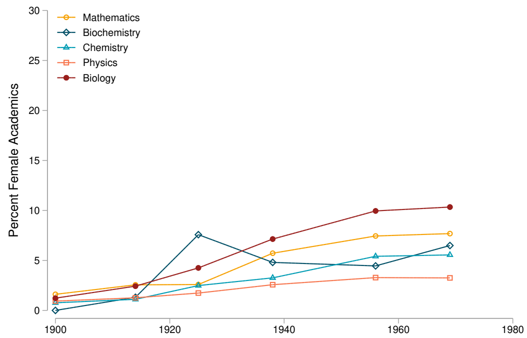
(a) *Sample 1:* All universities, all disciplines, 1900-1969



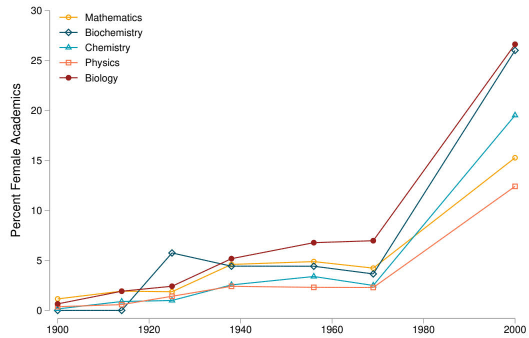
(b) *Sample 1:* All universities, all disciplines, 1900-1969



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



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



Notes: The Figure shows the percentage of female academics by discipline. Panel (a) shows female shares for all disciplines (aggregated at the level of Humanities, Social Sciences, and STEM) in all universities until 1969. Panel (b) shows female shares in nine exemplary disciplines in all universities until 1969. Appendix Figure Figure B.3 shows the remaining disciplines. Panel (c) shows female shares in the sciences (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Panel (d) shows female shares in the sciences in prestigious universities until 2000. The data were collected by the authors from various volumes of Minerva and department websites, see section 1 for details.

In the science sample (mathematics, physics, chemistry, biochemistry, and biology), female shares were lower than in many other disciplines but varied substantially across disciplines (Figure 5, panel c). In the sample of prestigious universities, female shares were even lower until 1969. In the last three decades of the 20th century, they increased substantially and by 2000 reached around 12% in physics, 15% in mathematics, 19% in chemistry, 26% in biochemistry, and 27% in biology (Figure 5, panel d).

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 output measures that are comparable across time and space. In contrast, comparable measures are usually not available in other occupations. Publications are key performance metrics that are used to evaluate potential hires, allocate research funds, and rank academics. As previously discussed, we do not interpret publications as the true ability of academics. They reflect gender differences in output that could stem from differences in preferences, discrimination in the peer-review process or in the workplace (e.g., because women had worse access to high-quality labs), and other gender imbalances (e.g., differences in childcare contributions).

3.1 Individual-Level Publication Gaps

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

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

where Pub_{it} measures the number of papers that scientist i from cohort $t(i)$, discipline $d(i)$, country $c(i)$, and university $u(i)$ published in journals covered by the Web of Science. As described above, we count papers in a $\pm$ five-year window around scientist i's cohort $t(i)$. 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 female indicator Female_i with indicators for four different periods: pre-WW1 (1900 and 1914 cohorts), interwar (1925 and 1938), post-WW2 (1956 and 1969), and modern (2000). All regressions control for discipline-specific measures of experience, computed as the number of times a scientist is observed in the data.[24] We estimate each regression three times, controlling

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

[24] We include variables that indicate the number of times the scientist has been observed by cohort t. E.g., a scientist observed in 1956 and in 1969 has two observations. For the first observation in 1956, the indicator corresponding to observing the scientist for the first time equals 1. For the second observation in

for increasingly finer fixed effects:

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

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

The 1900 and 1914 cohorts of female scientists published, on average, 1.2 fewer papers than men in the full sample of all universities. The 1925 and 1938 cohorts of female scientists published 1.7 fewer papers, and the 1956 and 1969 cohorts published 2.2 fewer papers (Table 2, sample 2, column 1, significant at the 1% level). These are substantial gaps compared to the mean of publications, which was around 4. Even comparing women to men within the same cohort and university (column 2), the publication gaps only shrink slightly. Note, however, that the university may be endogenous, akin to occupations in traditional Mincer regressions. Finally, in column 3, we control for cohort-discipline-university fixed effects. We thus estimate publication gaps for scientists in the same university, discipline, and cohort (e.g., Harvard biologists in 2000). Even within this restricted comparison group, we find a similar pattern of gender gaps in publications.

The coverage of journals in the Web of Science and the propensity to publish vary over time, across countries, and across disciplines. This affects comparisons of publication gaps because women are not equally distributed. E.g., many women entered the data in later periods and worked in the United States, i.e., periods and a country with higher average publications. We therefore show alternative specifications that use standardized publications as the dependent variable. We standardize the number of publications to have a mean of 0 and a standard deviation of 1 within each country, cohort, and discipline (e.g., biology in the United States in 1969). Using this dependent variable, we find a negative gender gap in publications of around 0.21 s.d. for the 1900 and 1914 cohorts.

1969, the indicator corresponding to observing the scientist for the second time equals 1. The indicators for observing the scientist a third, fourth, or fifth time are all zero in this example. We include separate experience indicators for each discipline. Analyses restricted to each scientist's first observation yield similar results (Table 3). This suggests that gender gaps in publications do not stem from observing women at different career stages than men.

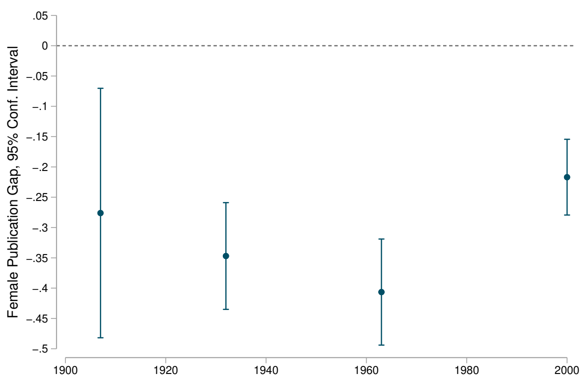
Table 2: Gender Gaps in Individual-Level Publications

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:		Publications			Standardized Publications	
Sample 2: All Universities 1900-1969						
Female (1900/14)	-1.169***	-0.600	-1.003**	-0.213**	-0.143	-0.222**
	(0.435)	(0.508)	(0.503)	(0.085)	(0.103)	(0.100)
Female (1925/38)	-1.700***	-1.249***	-1.611***	-0.246***	-0.195***	-0.215***
	(0.417)	(0.334)	(0.411)	(0.022)	(0.031)	(0.040)
Female (1956/69)	-2.210***	-1.263***	-1.413***	-0.254***	-0.154***	-0.162***
	(0.518)	(0.278)	(0.284)	(0.024)	(0.016)	(0.014)
Observations	67,618	67,618	67,618	67,618	67,618	67,618
R-squared	0.169	0.251	0.344	0.005	0.129	0.228
Sample 3: Prestigious Universities 1900-2000						
Female (1900/14)	-1.638***	-1.273**	-1.711***	-0.276***	-0.244**	-0.323***
	(0.595)	(0.556)	(0.565)	(0.105)	(0.106)	(0.111)
Female (1925/38)	-2.635***	-2.070***	-2.448***	-0.347***	-0.266***	-0.293***
	(0.627)	(0.528)	(0.645)	(0.045)	(0.047)	(0.058)
Female (1956/69)	-3.589***	-2.766***	-3.005***	-0.406***	-0.326***	-0.330***
	(0.750)	(0.582)	(0.654)	(0.045)	(0.037)	(0.029)
Female (2000)	-4.290***	-3.845***	-3.561***	-0.217***	-0.201***	-0.186***
	(0.782)	(0.669)	(0.631)	(0.032)	(0.029)	(0.027)
Observations	88,537	88,537	88,537	88,537	88,537	88,537
R-squared	0.245	0.256	0.279	0.019	0.058	0.113
Experience×Discipline	Yes	Yes	Yes	Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes		Yes	Yes	
Cohort×University FE		Yes			Yes	
Cohort×Discipline×University FE			Yes			Yes

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist-level. Sample 2 includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Sample 3 includes scientists in prestigious universities until 2000. In columns 1-3, the dependent variable 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 publications, standardized at the cohort-discipline-country level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman, interacted with the relevant cohort(s). The regressions also control for experience by discipline and different sets of fixed effects (see definition (2) for details). Standard errors are clustered at the discipline-country level, with 418 clusters in sample 2 and 183 in sample 3. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

The gender gap increased in absolute magnitude for the 1925 and 1938 cohorts and then slowly declined over the second half of the 20th century (Table 2, sample 2, columns 4-6).

Figure 6: Gender Gaps in Publications over Time



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

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.

Table 3: Individual-Level Publication Gaps (Robustness)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	±Three-Year Window		±Ten-Year Window		Unique Matches		Full Professors		First Cohort	
Dependent Variable	Publications	Standard. Publications	Publications	Standard. Publications	Publications	Standard. Publications	Publications	Standard. Publications	Publications	Standard. Publications
Sample 2: All Universities 1900-1969										
Female (1900/14)	-0.802**	-0.204**	-2.703***	-0.253***	-1.163***	-0.206**	-1.257***	-0.312***	-0.978**	-0.178*
	(0.312)	(0.081)	(0.715)	(0.059)	(0.441)	(0.086)	(0.286)	(0.031)	(0.465)	(0.095)
Female (1925/38)	-1.089***	-0.225***	-2.913***	-0.251***	-1.719***	-0.247***	-2.020***	-0.295***	-1.689***	-0.253***
	(0.263)	(0.024)	(0.692)	(0.025)	(0.428)	(0.024)	(0.583)	(0.038)	(0.451)	(0.025)
Female (1956/69)	-1.500***	-0.248***	-2.760***	-0.250***	-2.268***	-0.261***	-2.728***	-0.288***	-2.143***	-0.258***
	(0.351)	(0.023)	(0.630)	(0.022)	(0.510)	(0.024)	(0.605)	(0.041)	(0.457)	(0.023)
Observations	67,182	67,182	66,700	66,700	59,960	59,960	35,619	35,619	49,996	49,996
# Clusters (std. errors)	417	417	416	416	412	412	383	383	413	413
R-squared	0.162	0.005	0.176	0.005	0.169	0.008	0.203	0.017	0.171	0.010
Sample 3: Prestigious Universities 1900-2000										
Female (1900/14)	-1.143***	-0.265***	-3.487***	-0.316***	-1.621***	-0.271**	-1.541***	-0.350***	-1.492**	-0.235**
	(0.398)	(0.090)	(1.036)	(0.075)	(0.590)	(0.104)	(0.402)	(0.058)	(0.625)	(0.116)
Female (1925/38)	-1.669***	-0.309***	-4.408***	-0.352***	-2.642***	-0.346***	-2.914***	-0.404***	-2.538***	-0.350***
	(0.409)	(0.048)	(1.013)	(0.053)	(0.632)	(0.047)	(0.725)	(0.081)	(0.675)	(0.046)
Female (1956/69)	-2.355***	-0.387***	-4.426***	-0.407***	-3.522***	-0.403***	-4.475***	-0.451***	-3.416***	-0.411***
	(0.516)	(0.045)	(0.874)	(0.045)	(0.756)	(0.047)	(0.956)	(0.056)	(0.729)	(0.055)
Female (2000)	-2.753***	-0.204***	-7.486***	-0.228***	-4.344***	-0.223***	-0.966	-0.103***	-4.314***	-0.217***
	(0.501)	(0.032)	(1.371)	(0.033)	(0.777)	(0.031)	(0.771)	(0.030)	(0.792)	(0.032)
Observations	88,280	88,280	88,077	88,077	78,464	78,464	37,139	37,139	75,390	75,390
# Clusters (std. errors)	183	183	183	183	182	182	178	178	183	183
R-squared	0.242	0.017	0.259	0.019	0.239	0.022	0.309	0.039	0.243	0.021
Experience×Discipline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

Second, in this sample we observe nearly all universities in each of the seven cohorts. As a result, compositional changes in the sample of universities cannot affect the findings.

In this sample, the publication gaps are larger (in absolute terms) 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 0.28 s.d. for the 1900 and 1914 cohorts to around 0.41 s.d. for the 1956 and 1969 cohorts and then declined to 0.22 s.d. for the 2000 cohort (see also Figure 6).

Robustness

The estimated gender gaps in publications are robust to alternative ways of linking papers to academics. First, results are similar if we measure publications over a ± 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). Similarly, we show that results are robust if we measure publications over a ± ten-year window (Table 3, columns 3-4). We also show that publication gaps are very similar in a sample of scientists with unique surname - first initial - discipline combinations in every cohort (Table 3, columns 5-6). This suggests that gender differences in publications do not stem from gender differences in the frequency of certain surname - first initial pairs. Publication gaps are also similar in the sample of full professors, the academic rank that is most comparable across countries (Table 3, columns 7-8). 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 9-10). The estimated coefficients are very similar if we control for more stringent fixed effects, as defined in equation (2) (results unreported).

3.2 Linking Gender Gaps in Hiring and Publications

Figure 6 shows a downward and then upward-sloping pattern of gender gaps in publications over time. In contrast, we show in the first part of the paper that the share of women increased over the 20th century (Figure 2). Changes in the share of women in academia may be related to relative changes in the selection and publishing opportunities of men and women. To explore whether gender gaps in hiring and publishing are systematically linked, we propose and estimate a model along the lines of Roy (1951).

3.2.1. Model

The model allows for (i) selection on unobservables in the hiring market, (ii) gender bias in hiring, and (iii) gender bias in the publication market. These factors contribute to the gender gap in publications through (a) *indirect* effects of selection and gender bias in the hiring market, and (b) *direct* effects of gender bias in the publication market.[25]

At the hiring stage, denoted by 0, any academic position i can be filled with a woman W or a man M. Women and men face differential barriers until they are hired as academics. Such barriers may be institutional, e.g., certain high 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). Barriers may also stem from gender differences in exposure to academic role models (e.g., Bell et al. 2019). We refer to any such gender bias in hiring as Δ_0. We express selection in the hiring market in terms of s_0^W, the share of women among all hired scientists. During the expansion of the university sector, the total number of male and female academics increased dramatically. Furthermore, the *share* of women increased over time because relatively more women were hired (Figure 2).

At a later stage, denoted by 1, we observe publication outcomes for all hired scientists. We refer to any gender bias in publications as Δ_1. The model allows for the possibility that selection in hiring affects observed publications, 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 simple version of the model with the following assumptions:

[25] A more general model also incorporating gender gaps in citations (section 4) and promotions (section 5) would have to impose too many (and controversial) assumptions to remain tractable. For example, the effect of selection in publishing on citations is unclear from a modeling perspective. It could be positive for some women and negative for others. Our results indicate that women published fewer papers and, hence, some women only published their highest-quality ideas. Other women, instead, may have been impeded from carrying out high-quality research, e.g., because of a disproportional amount of housework, which may have resulted in fewer and lower-quality publications. A model that combines these two countervailing forces would have to rely on unreasonable assumptions to remain tractable.

(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, we present a more general version of the model that relaxes assumption (ii) and other parametric assumptions (assumptions (iii)-(iv) below).

Selection in the Hiring Market

Suppose that academic position i can 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 is 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 everything else fixed, when $\Delta_0 > 0$ women need to overcome the additional hurdle or gender bias Δ_0 to be hired. Assuming that $\epsilon_{0i} \equiv \epsilon_{0i}^W - \epsilon_{0i}^M$ is i.i.d. standard normal (assumption (iv) below), the probability that a woman is hired is:

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

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

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

The differences in observable characteristics X_i could be SAT scores, college GPA, the specialization of the undergraduate degree, or differential treatment of boys and girls while growing up. Such data are not available at a worldwide scale over the 20th century and, if available, they would be affected by various sources of selectivity and measurement error. Therefore, we assume $X_i = 0$, i.e., men and women are a priori equally qualified for the academic position, 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 scientists. In this case, as we can directly measure s_0^W in the data, we compute Δ_0 without needing to perform any estimation.[26]

Publication Market

Conditional on academic position i being filled by either a woman or a man, we observe

[26] 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 in equation (6).

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. I.e., if academic position i is filled by a woman, we observe the publication outcome of a woman, otherwise a man's outcome. Since for any i we cannot observe the counterfactual publication outcome (i.e., the publications if position i had been filled by the other gender), equation (8) will be subject to selection on unobservables if the error terms in (5) and (8) are correlated. Δ_1 is the gender bias in publications, which may reflect gender imbalances in working conditions, preferences, or other constraints that differentially affected publications, such as discrimination in the peer-review process, in the workplace, or other gender imbalances (e.g., differences in childcare responsibilities). In the 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 i.i.d. standard normal.

These assumptions are not necessary for identification but simplify estimation.[27] 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. $\rho_W > 0$ represents positive selection of women. Similarly, because $\epsilon_{0i} \equiv \epsilon_{0i}^W - \epsilon_{0i}^M$, $\rho_M < 0$ represents positive selection of men.

Publication Outcome Conditional on Gender

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

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

where $\lambda(\cdot) \equiv \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio.[28] Analogously, the expectation of Y_{1i}^M conditional on $Y_{0i} \leq 0$ is:

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

For men, when the gender bias in publications Δ_1 is large, the conditional expectation of publications Y_{1i}^M is large. Moreover, equation (10) indicates that when the gender bias in hiring Δ_0 is large (Figure 2), as $\lambda\left(\Delta_0\right)$ is close to zero, the conditional expectation of Y_{1i}^M may be unaffected, even if $\rho_M \neq 0$. For women, when the gender bias in hiring Δ_0 is large,

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

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

the inverse Mills ratio $\lambda(-\Delta_0)$ is high, and the conditional expectation of publications Y_{1i}^W is large if $\rho_W > 0$. This holds independently of any gender bias Δ_1 in publications.

We combine equations (9) and (10) and assume that we observe the same characteristics (e.g., cohort, university, discipline) whether position i is filled by a woman or a man $Z_i^W = Z_i^M = Z_i$, to obtain an expression for the gender gap function in publications:[29]

$$\begin{aligned}
&\mathbb{E}\left[Y_{1i}^M \middle| Z_i, Y_{0i} \le 0\right] + \text{Female}_i \times \left[\mathbb{E}\left[Y_{1i}^W \middle| Z_i, Y_{0i} > 0\right] - \mathbb{E}\left[Y_{1i}^M \middle| Z_i, Y_{0i} \le 0\right]\right] \\
&= Z_i\beta_1 + \Delta_1 - \rho_M\,\lambda(\Delta_0) + \text{Female}_i \times \left[\rho_W\,\lambda(-\Delta_0) - \Delta_1 + \rho_M\,\lambda(\Delta_0)\right] \qquad (11)\\
&= Z_i\beta_1 + \Delta_1 - \rho_M\,\lambda\left(-\Phi^{-1}\left(s_0^W\right)\right) + \text{Female}_i \times g\left(s_0^W\right),
\end{aligned}$$

where the Female_i indicator denotes whether scientist i is a woman and $g\left(s_0^W\right) \equiv \rho_W\,\lambda\left(\Phi^{-1}\left(s_0^W\right)\right) - \Delta_1 + \rho_M\,\lambda\left(-\Phi^{-1}\left(s_0^W\right)\right)$ is the *gender gap function* in publications which depends on the share of women in the profession s_0.[30] Equation (11) highlights how, in addition to any direct gender bias in publications (Δ_1), observed gender gaps in publications can be indirectly affected by gender biases in hiring (Δ_0).

3.2.2. Estimation Results: "Gender U"

We estimate equation (11) allowing the gender gap function to vary by period $p = 1, 2, 3$:

$$\begin{aligned}
\text{Pub}_{it} \;=\;& \delta + \text{Female}_i \times 1[t(i) = 1900-38] \times \Bigg[\underbrace{\rho_1^W\,\lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \Delta_{11} + \rho_1^M\,\lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)}_{g_1\left(s_{0\ell(i)}^W\right) \equiv \text{ gender gap function in 1900-38}}\Bigg] \\
&+ \text{Female}_i \times 1[t(i) = 1956/69] \times \Bigg[\underbrace{\rho_2^W\,\lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \Delta_{12} + \rho_2^M\,\lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)}_{g_2\left(s_{0\ell(i)}^W\right) \equiv \text{ gender gap function in 1956/69}}\Bigg] \\
&+ \text{Female}_i \times 1[t(i) = 2000] \times \Bigg[\underbrace{\rho_3^W\,\lambda\left(\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right) - \Delta_{13} + \rho_3^M\,\lambda\left(-\Phi^{-1}\left(s_{0\ell(i)}^W\right)\right)}_{g_3\left(s_{0\ell(i)}^W\right) \equiv \text{ gender gap function in 2000}}\Bigg] \\
&+ \text{Experience}_{it}\boldsymbol{\delta}_{\text{exp}} + \text{FE}(i,t) + \varepsilon_{it}, \qquad (12)
\end{aligned}$$

where Pub_{it} measures the standardized number of papers published by scientist i in cohort $t(i)$, $s_{0\ell(i)}^W$ is the share of female scientists in i's cohort-country $\ell(i)$ (e.g., the United States in 2000).[31,32] We control for the term $Z_i\beta_1 + \Delta_1 - \rho_M\,\lambda\left(-\Phi^{-1}\left(s_0^W\right)\right)$ in equation

[29] We assume $Z_i^W = Z_i^M = Z_i$ to control for these characteristics by fixed effects. If more granular data on (Z_i^W, Z_i^M) were available, $Z_i\beta_1$ in equation (11) would be replaced by $(1 - \text{Female}_i)Z_i^M\beta_1 + \text{Female}_i Z_i^W\beta_1$.

[30] The second equality in (11) follows from equation (7) when $X_i = 0$, $\Phi^{-1}\left(s_0^W\right) = -\Delta_0$.

[31] For cohort-country pairs with $s_{0\ell}^W$ equal to zero, $\Delta_0 = -\Phi^{-1}\left(s_{0\ell}^W\right)$ is not defined. Hence, the regression excludes observations from cohort-country pairs that do not contain women. In contrast, the model in Appendix C is defined for cohort-country pairs with $s_{0\ell}^W$ equal to zero. To ensure comparability, we use the same sample for all columns of Table 4. However, the estimation results of regression (C.8) are robust to including observations from cohort-country pairs that do not include any woman.

[32] To reduce measurement error in the regressors $\lambda\left(\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ and $\lambda\left(-\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ due to very low shares of female scientists in some cohorts, disciplines, and countries, we compute the share of female scientists $s_{0\ell}^W$ at the level of the cohort-country ℓ, rather than at a finer level the cohort-discipline-country.

(11) with cohort-discipline-country (or finer) fixed effects (see definition (2) for details) and experience indicators.[33] While the fixed effects in regression (12) capture the average number of publications among male scientists, the gender gap functions capture any systematic difference in the publications of women versus men.

The gender gap function $g_p\left(s_{0\ell}^W\right)$ includes the difference between two gender-specific inverse Mills ratios. While each individual inverse Mills ratio is strictly decreasing in its argument, their difference does not need to be decreasing. Importantly, the model does not "force" any specific shape for $g_p\left(s_{0\ell}^W\right)$ on the data, in that a lack of gender gaps, an increasing or decreasing relationship, a straight line, a U, or an inverted U could all be estimated. The estimated parameters $\left(\widehat{-\Delta_{1p}}, \widehat{\rho}_p^W, \widehat{\rho}_p^M, p=1,2,3\right)$ determine the shape of the gender gap as a function of the share of female scientists.

Table 4: Individual-Level Publication Gaps and the Share of Females

Dependent Variable	(1) Standard. Publications	(2) Standard. Publications	(3) Standard. Publications		(4) Standard. Publications	(5) Standard. Publications	(6) Standard. Publications
	Inverse Mills Ratios, regression (12)				*Polynomials, regression (C.8)*		
Female (1900/38)							
$-\Delta_{11}$	-5.117**	-4.526**	-5.379**	γ_{01}	0.010	-0.012	0.002
	(1.970)	(2.033)	(2.142)		(0.159)	(0.182)	(0.213)
ρ_1^W	1.833**	1.620**	1.926**	γ_{11}	-12.645**	-9.415	-11.329
	(0.718)	(0.744)	(0.784)		(5.373)	(6.059)	(6.991)
ρ_1^M	8.540*	7.835*	9.398*	γ_{21}	90.494*	67.697	81.963
	(4.510)	(4.660)	(4.911)		(50.049)	(51.859)	(56.222)
Female (1956/69)							
$-\Delta_{12}$	-5.504***	-3.747**	-2.577	γ_{02}	-0.031	-0.017	-0.145
	(1.854)	(1.820)	(1.832)		(0.178)	(0.174)	(0.173)
ρ_2^W	2.063***	1.404*	0.923	γ_{12}	-11.203**	-8.824*	-5.193
	(0.745)	(0.735)	(0.742)		(4.898)	(4.862)	(4.785)
ρ_2^M	7.327***	4.527*	2.971	γ_{22}	63.204**	45.212	25.827
	(2.770)	(2.714)	(2.747)		(28.500)	(27.636)	(27.204)
Female (2000)							
$-\Delta_{13}$	-2.173**	-1.670*	-1.293	γ_{03}	0.220	0.192	0.175
	(0.977)	(0.912)	(0.875)		(0.133)	(0.121)	(0.109)
ρ_3^W	1.053**	0.813*	0.634	γ_{13}	-3.342***	-2.931***	-2.636***
	(0.491)	(0.457)	(0.437)		(0.957)	(0.880)	(0.808)
ρ_3^M	1.320	0.908	0.597	γ_{23}	5.158***	4.344***	3.736***
	(0.809)	(0.761)	(0.737)		(1.600)	(1.504)	(1.411)
Observations	82,674	82,674	82,674		82,674	82,674	82,674
R-squared	0.018	0.055	0.107		0.018	0.055	0.107
Experience×Discipline	Yes	Yes	Yes		Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes			Yes	Yes	
Cohort×University FE		Yes				Yes	
Cohort×Discipline×University FE			Yes				Yes

Notes: The Table shows estimation results for equation (12) in columns 1-3, and for its more general version, equation (C.8), in columns 4-6. The sample includes scientists (mathematics, physics, chemistry, biochemistry and biology) in prestigious universities until 2000 (sample 3). The dependent variable equals the standardized number of publications in a ± five-year window around a cohort (i.e., 1995-2005 for a scientist listed in 2000). In columns 1-3, the main explanatory variables are indicators that equal 1 if the scientist is a woman, interacted 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 computed within sample 3. In columns 4-6, the main explanatory variables are indicators that equal 1 if the scientist is a woman, interacted with second-degree polynomials of $s_{0\ell}^W$. All regressions exclude observations from cohort-country combinations that do not include any woman. The regressions also control for experience by discipline and different sets of fixed effects (see definition (2) for details). Standard errors are clustered at the discipline-country level, with 174 clusters. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

[33]This assumes that Z_i varies at most at the level of the fixed effects included in regression (12). In addition, because we measure s_0^W at the cohort-country level, the fixed effects in regression (12) always fully control for the occurrence of $\rho_M\,\lambda\left(-\Phi^{-1}\left(s_0^W\right)\right)$ *not* interacted with Female$_i$ in equation (11).

We report estimates of regression (12) for sample 3 in Table 4, columns (1)-(3). We estimate a negative constant component of the gender bias in publications ($\widehat{-\Delta}_{1p}$, for each period $p = 1, 2, 3$). This suggests that women faced hurdles in publishing that were independent of the share of female scientists in any country and cohort. Such hurdles could, for example, be due to higher levels of childcare responsibilities of academic mothers relative to fathers (e.g., Moser and Kim 2021). The constant gender bias in publications is smaller in 2000 than in the past (-2.17 versus around -5 in Table 4, column (1)). This is consistent with, for example, the increased availability of professional childcare that enabled academic mothers to devote more time to publishing while having small children.

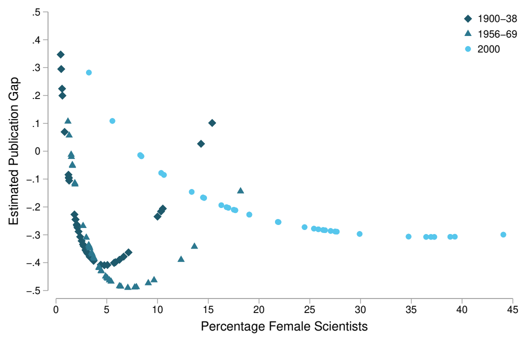
Using the estimated $\widehat{-\Delta}_{1p}$, as well as $\widehat{\rho}_p^W$ and $\widehat{\rho}_p^M$, we compute the predicted gender gap function $\widehat{g}_p\left(s_{0\ell}^W\right) = \widehat{\rho}_p^W \lambda\left(\Phi^{-1}\left(s_{0\ell}^W\right)\right) \widehat{-\Delta}_{1p} + \widehat{\rho}_p^M \lambda\left(-\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ and plot it against the share of female scientists in each cohort-country, separately for each period p (Figure 7, panel a). Each dot in the figure represents the predicted gender gap in publications for period p as a function of the share of women scientists in each cohort-country pair (e.g., the United States in 2000). The figure suggests a U-shaped relationship between the gender gap in publications and the share of female scientists. We refer to this relationship, in short, as the "gender U." The gender U arises because the estimates $\widehat{\rho}_p^W$ and $\widehat{\rho}_p^M$ are both positive. Hence, with a rising share of women $s_{0\ell}^W$, $\widehat{\rho}_p^W \lambda\left(\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ moves from being large toward zero, while $\widehat{\rho}_p^M \lambda\left(-\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ moves in the opposite direction. This leads to small or even positive publication gaps in cohort-countries with very low shares of female scientists, the "Marie-Curie" cohort-countries. However, with increasing shares of women, gender gaps in publications become more negative. When the share of women increases beyond very low levels, the negative gender gap in publications decreases.

For the interpretation of the gender U, it is important to note that the number of publications only partly reflects the ability of academics but also reflects discrimination in the peer-review process, uneven allocation of resources (e.g., access to labs and research grants), and asymmetries in coauthorship networks. As a result, $\widehat{\rho}_p^W$ and $\widehat{\rho}_p^M$ partly capture selection in terms of "true" ability but also changes in discrimination and in essential publication inputs. Hence, the gender U can arise from different economic channels. For example, the downward-sloping part of the gender U can arise either because of positive selection of women or, probably less plausibly, because a higher female share was associated with more discrimination against women in the publication market. Both of these channels would increase the publication gap as a function of the female share. Similarly, the upward-sloping part of the gender U can arise, for example, due to negative selection of men, increasing opportunities for women to coauthor with other women, or rising empowerment of women, which could have led to reduced discrimination in the publication market and in the allocation of resources. These channels would all contribute to a closing publication gap as a function of the female share.

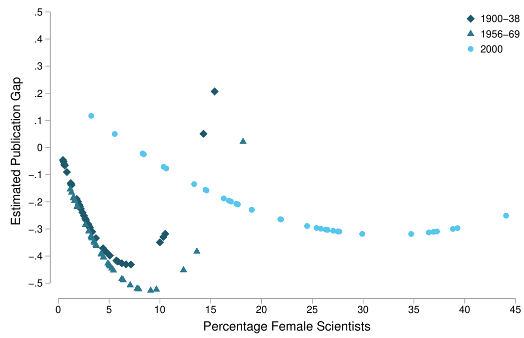
A plausible interpretation for the gender U is that the most talented women entered academia first (*selection effect*) and that a higher representation of women in academia

Figure 7: Gender Gaps in Publications and the Share of Women: *Sample 3*

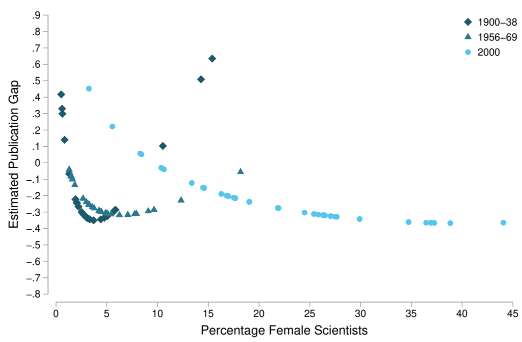
(a) All: Inverse Mill Ratios (IMRs)



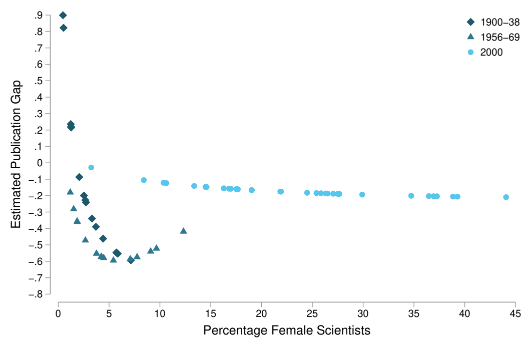
(b) All: Polynomials



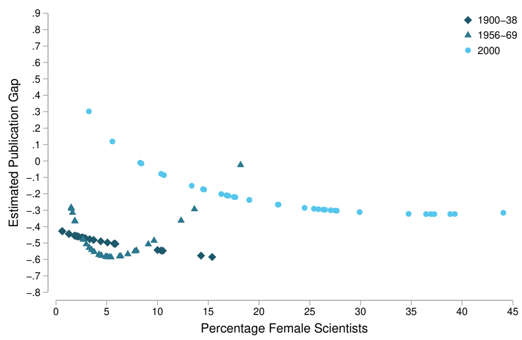
(c) Biology: IMRs



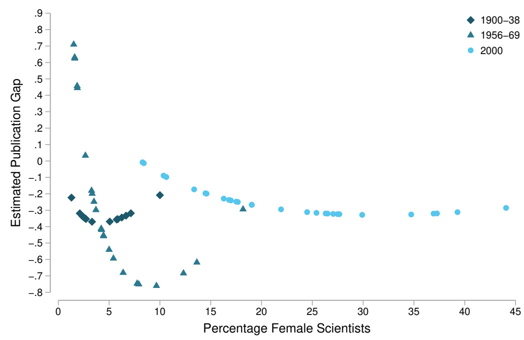
(d) Physics: IMRs



(e) Biochem. and Chemistry: IMRs



(f) Mathematics: IMRs



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, in all science disciplines (panels a-b) and separately by discipline (panels c-f), using the data from sample 3. Panel (a) plots estimated gender gap functions from regression (12), while panel (b) plots gender gaps from regression (C.8). Each dot in panels (a)-(b) represents the estimated gender gap in publications for a cohort-country as a function of the share of women scientists in that cohort-country, e.g., 2000 in the United States. Each of the panels (c)-(f) reports estimated gender gaps for a specific discipline. The discipline-specific gender gap functions in panels (c)-(f) are obtained by including discipline-specific interactions in regression (12). Each dot in panels (c)-(f) represents the gender gap in publications for a specific discipline in a cohort-discipline-country as a function of the share of women scientists in that country-cohort. All regressions control for experience-discipline and cohort-discipline-country fixed effects.

was accompanied by more publishing opportunities for women (which we refer to as *empowerment effect*).[34]

[34]The alternative explanation that the gender gap in publications widens at first due to an increase in discrimination against women in the publication market and then closes again due to negative selection of men strikes us as less plausible for two reasons. First, with rising female shares, increasing discrimination against women in the publication market is inconsistent with the evidence by Card et al. (2022) and Card et al. (2023), who find increasing *recognition* of female academics over time as measured by elections to prestigious scientific societies. Second, negative selection of men would imply that the best male scientists left academia as a result of female entry.

For low female shares, the selection effect dominates as the inverse Mills ratio $\widehat{\rho}_p^W \lambda\left(\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ is large and positive for female shares close to zero.[35] Additionally, for low female shares, the empowerment effect is weak as $\widehat{\rho}_p^M \lambda\left(-\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ is close to zero. This could happen because, with low female shares, it is harder for women to find female coauthors[36] and because only very few women had become editors or referees at prestigious journals. Once the female share rises beyond very low levels, the empowerment effect becomes more important as $\widehat{\rho}_p^M \lambda\left(-\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ increases. At the same time, the strength of the selection effect decreases as $\widehat{\rho}_p^W \lambda\left(\Phi^{-1}\left(s_{0\ell}^W\right)\right)$ shrinks. Hence, the gender gap in publications shrinks.

It is beyond the scope of one paper to provide a comprehensive analysis of the mechanisms behind the selection and empowerment effects at a global scale throughout the 20th century. However, the estimated gender U succinctly captures which of the opposing effects dominated in different periods and countries during the 20th century.

Robustness

The finding of a U-shaped relationship between the gender gap in publications and the share of female scientists is robust to estimating a more general version of the model, to using different samples of scientists, and also to estimating the gender gap function separately for each discipline.

First, we show estimates of the more general model outlined in Appendix C. This model relaxes the functional form restrictions embedded in the inverse Mills ratios by approximating the gender gap function in publications by second-degree polynomials of $s_{0\ell}^W$, i.e., estimating $g_p\left(s_{0\ell}^W\right) = \gamma_{0p} + \gamma_{1p} s_{0\ell}^W + \gamma_{2p}\left(s_{0\ell}^W\right)^2$. The results are reported in columns (4)-(6) of Table 4 and panel (b) of Figure 7.[37] The more general model confirms that the gender U is not an artifact of the functional forms embedded in regression (12).

Second, we also confirm the gender U for most periods and disciplines by estimating discipline-specific gender gap functions in regression (12) (Figure 7, panels c-f).[38] Third, we explore whether the gender U is a phenomenon restricted to prestigious universities or whether it also holds for the sample of all universities in all countries. We estimate

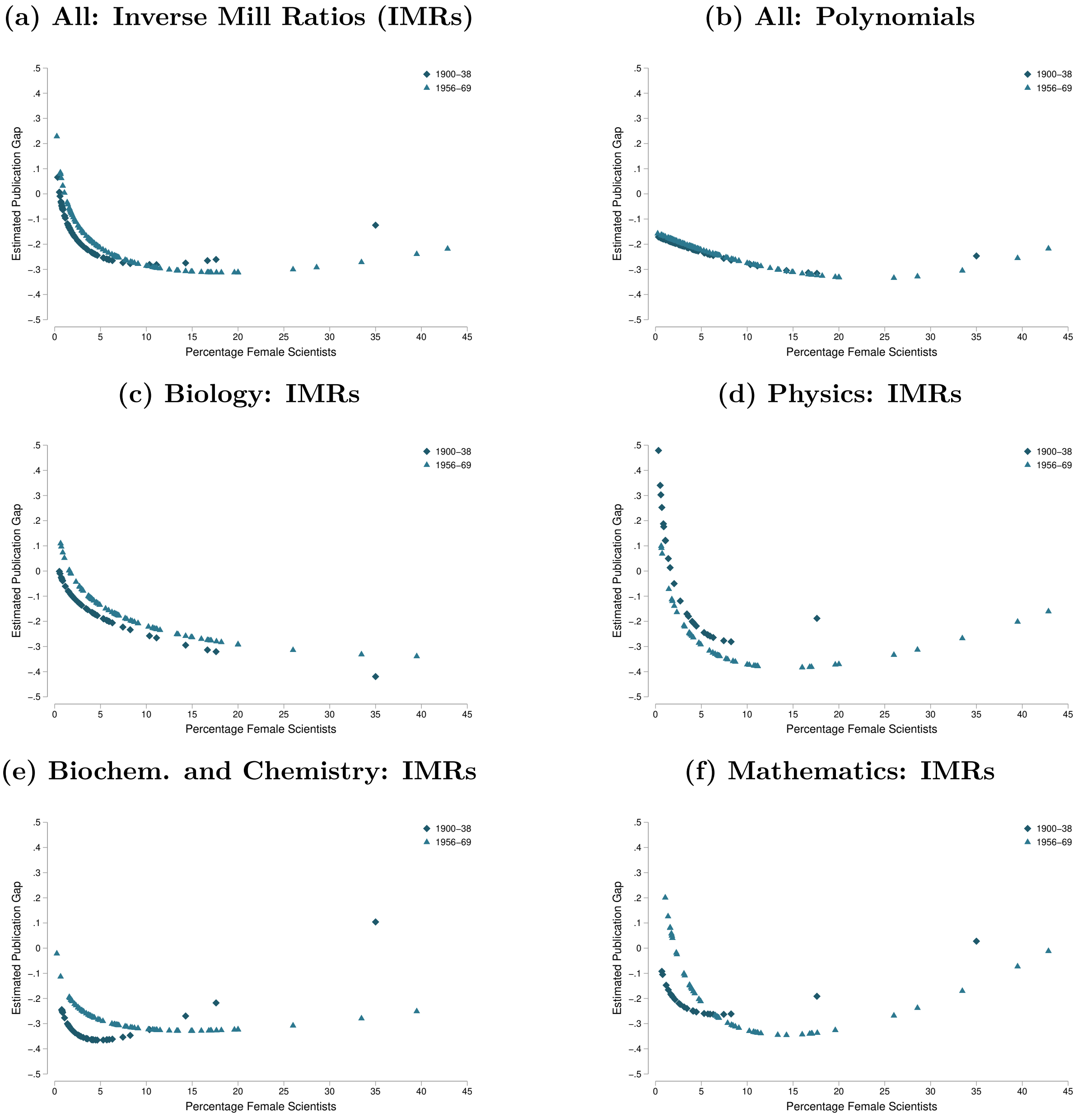
[35] This is also in line with the findings of Ashraf et al. (2022), who document positive selection of women into working for a large multinational firm for the period 2015-2019 and with Mulligan and Rubinstein (2008) who document positive selection of women in terms of ability in the general labor market.

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

[37] The estimates $\widehat{\gamma}_{0p}$, $p = 1, 2, 3$, in columns (4)-(6) of Table 4, do *not* represent estimates of the parameters $-\Delta_{1p}$, $p = 1, 2, 3$. The more general specification of the model in regression (C.8) relaxes assumption (ii) above, so that the gender bias in publications Δ_1 is allowed to be a function of the share of female scientists s_0^W. As a consequence, the gender bias in publications cannot be separately identified from the rest of the gender gap function in regression (C.8). In particular, for any given value of the gender gap function $g_p\left(s_{0\ell}^W\right)$, the parameter $-\Delta_{1p}$ can be equal to γ_{0p} only in the very special circumstance that $\rho_p^W \lambda\left(\Phi^{-1}\left(s_{0\ell}^W\right)\right) + \rho_p^M \lambda\left(-\Phi^{-1}\left(s_{0\ell}^W\right)\right) = \gamma_{1p} s_{0\ell}^W + \gamma_{2p}\left(s_{0\ell}^W\right)^2$. Similarly, the estimates $\left(\widehat{\gamma}_{1p}, \widehat{\gamma}_{2p}\right)$ do not represent estimates of the parameters $\left(\rho_p^W, \rho_p^M\right)$, $p = 1, 2, 3$.

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

Figure 8: Gender Gaps in Publications and the Share of Women: *Sample 2*



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, in all science disciplines (panels a-b) and separately by discipline (panels c-f), using the data from sample 2. Panel (a) plots estimated gender gap functions from regression (12), while panel (b) plots gender gaps from regression (C.8). Each dot in panels (a)-(b) represents the estimated gender gap in publications for a cohort-country as a function of the share of women scientists in that cohort-country, e.g., 2000 in the United States. Each of the panels (c)-(f) reports gender gaps for a specific discipline. The discipline-specific gender gap functions in panels (c)-(f) are obtained by including discipline-specific interactions in regression (12). Each dot in panels (c)-(f) represents the gender gap in publications for a specific discipline in a cohort-country as a function of the share of women scientists in that country-cohort. All regressions control for experience-discipline and cohort-discipline-country fixed effects.

regressions (12) and (C.8) for sample 2 and confirm a U-shaped relationship for both periods 1900-1938 and 1956-1969. We also confirm the gender U in all disciplines and periods for sample 2, with the exception of biology (Figure 8).

Overall, these results corroborate the robustness of the gender U, which is a general pattern that we observe across samples, disciplines, and over time.

4 Gender Gaps in Citations

In this section, we explore whether papers written by women received fewer citations. Gender gaps in citations shed light on differences in peer recognition of female-authored papers. We conduct this analysis at the paper level to abstract from the gender differences in publishing documented in the previous section.

4.1 A Novel Procedure to Predict Citations

A key challenge when estimating gender gaps in citations is that men and women may work on different topics with different citation potentials. We control for differences in citation potential using a newly developed machine learning approach that uses paper titles to predict each paper's expected number of citations (see Appendix D for details).[39] A similar approach can be used to study performance or pay gaps between different groups with data on, e.g., occupational task descriptions, full text of job advertisements, or performance reviews.

We first filter all non-alphanumerical characters from papers' titles, remove common words (stopwords, e.g., "the"), and stem the words. Next, we extract all unigrams (i.e., words) and bigrams (i.e., two-word combinations) from the title of each of the N papers to obtain a paper-1,2-gram matrix $\boldsymbol{X}$ with entries x_{pj}, where p denotes papers and j denotes unigrams and bigrams.[40] As is common in text-based machine learning, we then reweight the matrix using term-frequency inverse-document frequency (tf-idf) reweighting. This decreases the relative importance of n-grams that carry little information but appear in many papers, for example, "study" or "method." The unigrams and bigrams then form the input for an L2-regularized regression model (ridge regression), which minimizes:

$$\min_{(\omega_j)_{j=1}^{W}} \left\{ \sum_{p=1}^{N} \left(y_p - \sum_{j=1}^{W} \omega_j \cdot x_{pj} \right)^2 + \lambda \sum_{j=1}^{W} \omega_j^2 \right\}, \tag{13}$$

where y_p are the total citations of paper p (standardized by country, discipline, and cohort). To reduce the importance of outliers, we winsorize citations at the 99th percentile (by discipline and cohort).[41] The main explanatory variables are the W indicators of the unigrams and bigrams that correspond to the respective entries of the paper-1,2-gram matrix $\boldsymbol{X}$. We additionally include indicators for the number of words in paper p's title. To allow for differences in citation patterns over time and across disciplines, we train separate models for each of the five scientific disciplines in each of the seven cohorts. For each discipline and cohort, we choose the optimal normalization strength λ using 10-fold cross-validation. The algorithm predicts standardized citations $\hat{y}_p$ for each paper p.

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

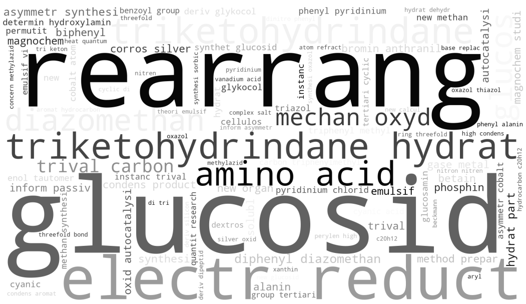
[40]Importantly, the Web of Science translates almost all titles into English.

[41]The results are very similar if we do not winsorize citations (Appendix Table D.1)

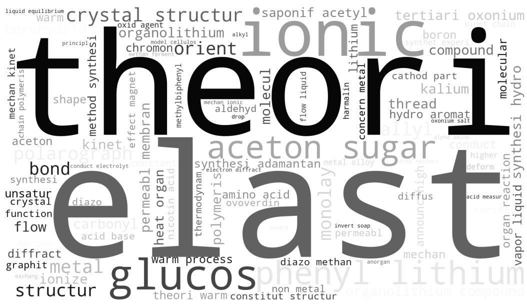
We use two approaches to estimate equation (13). For the first approach, the training sample consists of the universe of papers published by all scientists in our data. For the second approach, the training sample solely consists of papers published by men, predicting the actual citations of each paper as if it had been published by men. A model trained on all papers may give a better prediction of the actual citations. In contrast, a model solely trained on papers by men would address the concern that, for any given paper title, citations of women may be downward biased because of discrimination (see Barocas and Selbst (2016) for an overview of machine learning biases).

Figure 9: Words that Predict High Citations in Chemistry over Time

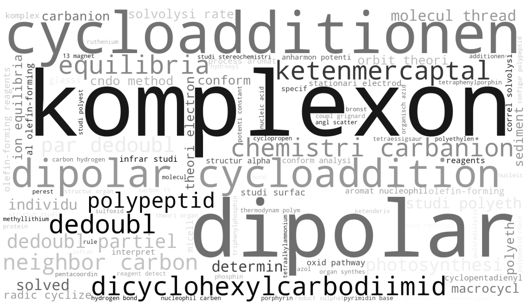
(a) 1900-1914



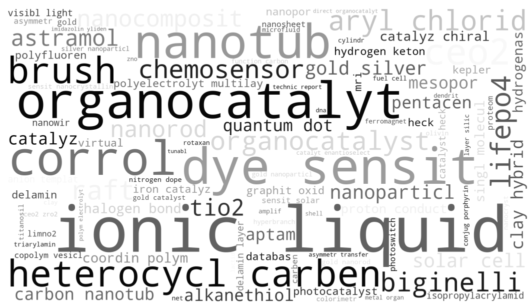
(b) 1925-1938



(c) 1965-1969



(d) 2000



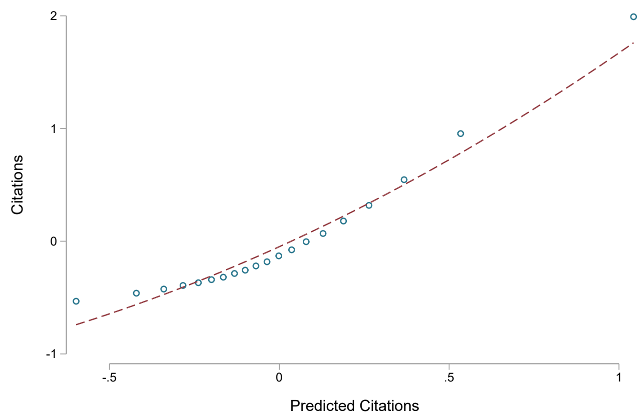
Notes: The Figure shows the unigrams and bigrams that predict the highest citations in chemistry for the indicated cohorts. The n-grams are identified with an L2-regularized regression model (ridge regression) that uses unigrams and bigrams of the title as inputs, see section 4.1 for details. A very 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 9 summarizes the unigrams and bigrams that predict high citations for chemistry and how they evolved over time.[42] For example, for the 1900 and 1914 cohorts, the classifier detects the names of sugar molecules ("glucos", "glucosid") whose chemical structures were described in the 1890s by the chemistry Nobel laureate Emil Fischer. Another highly cited n-grams in the same cohort is "triketohydrindane" (an alternative name for 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, Nylon by the DuPont chemist Wallace Carothers in 1935 and Perlon

[42] To save space, we show word clouds for the model trained on papers by male scientists and combine up to two cohorts in Figure 9. However, for the regression results reported below, we implement the prediction at the cohort-discipline level.

by the I.G. Farben chemist Paul Schlack in 1938. For the 1956 and 1969 cohorts, the classifier detects “dicyclohexylcarbodiimid”, which is a frequently used coupling agent for peptide synthesis based on the work by John Sheehan and George Hess in 1955. Another notable n-gram is “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 the Nobel laureate John Pople. Finally, for the 2000 cohort, the classifier detects “LiFePo4,” the chemical formula for lithium iron phosphate — a cathode material that is used for batteries which was discovered in 1996. Another stem for the 2000 cohort is “organocatalyt,” showing the importance of organocatalysis for which Benjamin List and David Macmillan shared the Nobel Prize in 2021.

Figure 10: Predicted and Actual Citations



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

In addition to identifying intuitive relationships between words and citations, the model performs well in predicting a paper’s actual citations. Figure 10 visualizes the strong positive relationship between predicted and actual citations ($R^2 = 0.31$).[43] As the figure suggests, the R^2 increases further when we include a second-order polynomial of the predicted citations. In contrast, higher-order polynomials do not lead to further increases in the R^2. We, therefore, control for the first and second-degree polynomials of predicted citations, which we interact with discipline indicators in our baseline regressions.[44]

4.2 Paper-Level Citation Gaps

We estimate citation gaps at the paper level, depending on whether papers were published by men or women. Importantly, we add our novel measures of predicted citations as a regressor to control for the fact that women may work on topics with less citation potential

[43] 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).

[44] Results are very similar if we control for predicted citations either linearly or non-parametrically (Appendix Table D.1).

than men. We estimate the following paper-level regression:

$$
\begin{aligned}
\text{Citations}_{pt} &= \gamma_1 + \gamma_2 \text{Female}_p \times 1[t(p) = 1900/14] + \gamma_3 \text{Female}_p \times 1[t(p) = 1925/38] \\
&+ \gamma_4 \text{Female}_p \times 1[t(p) = 1956/69] + \gamma_5 \text{Female}_p \times 1[t(p) = 2000] \\
&+ \widehat{\text{Citations}}_{pt} \boldsymbol{\gamma}_6 + \text{FE}(p,t) + \xi_{pt},
\end{aligned} \tag{14}
$$

where p denotes a paper and $t(p)$ the cohort in which the paper was published. The dependent variable is the number of standardized citations of paper p. In many scientific disciplines, authors are not ordered alphabetically. Instead, the first author is the one who conducts most of the research, while the last author is usually the most senior scientist who supervises the project. We, therefore, define an indicator (Female_p) that equals one if either the first or the last author of paper p is female.[45] The main explanatory variables are the interactions of Female_p with indicators for four time periods: pre-WW1 (1900 and 1914 cohorts), interwar (1925 and 1938), post-WW2 (1956 and 1969), and modern (2000).

Importantly, we control for the first and second-degree polynomials of predicted citations of paper p, $\widehat{\text{Citations}}_p$, interacted with discipline indicators.[46] We also control for various sets of fixed effects defined at the paper-level, adapting definition (2) accordingly.[47] Note that universities, disciplines, titles, and hence predicted citations are potentially endogenous. Therefore, the estimates should be interpreted as a decomposition of citation gaps into a part that can be explained by these factors and into an unexplained part due to other biases. We cluster standard errors at the discipline-country level.

Papers published by female scientists from the 1900-1914 cohorts received 0.12 s.d. fewer citations than papers published by male scientists in the sample of all universities (sample 2). Citation gaps were 0.15 s.d. for the 1925 and 1938 cohorts and 0.13 s.d. for the 1956 and 1969 cohorts (Table 5, sample 2, column 1).

[45] In mathematics, authors are mostly ordered alphabetically. As many mathematics papers are written by only one or two authors, this definition correctly captures the authors' gender for these papers. We show that results are very similar for alternative definitions of female-authored papers (Table D.2).

[46] In a recent paper, Koffi (2021) proposes a different method to estimate whether papers by female authors are under-cited. The method uses text similarity to identify papers that should have been cited. Our method, in contrast, controls for differences in the citation potential of papers. Our method estimates overall citation gaps, not just omissions of citations among the most similar papers. Furthermore, our method can be used when female shares are low and, hence, papers by women are unlikely to be among the most similar papers. Lastly, our methodology can also be used if the number of papers is large, due to its larger computational efficiency. Koffi's approach requires the calculation of pairwise similarities between all papers. Hence, the necessary calculations grow quadratically in the number of papers.

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

Table 5: Gender Gaps in Citations: Controlling for Predicted Citations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	Standardized Citations						
Sample 2: All Universities, Sciences, 1900-1969							
Female-First/Last Author (1900/14)	-0.118	-0.143*	-0.077*	-0.159**	-0.075**	-0.188***	-0.117***
	(0.105)	(0.082)	(0.042)	(0.070)	(0.033)	(0.068)	(0.034)
		[0.132]	[0.055]	[0.122]	[0.062]	[0.126]	[0.071]
Female-First/Last Author (1925/38)	-0.153***	-0.130***	-0.077***	-0.100**	-0.084**	-0.095**	-0.091**
	(0.045)	(0.035)	(0.027)	(0.040)	(0.036)	(0.038)	(0.040)
		[0.051]	[0.034]	[0.062]	[0.042]	[0.062]	[0.044]
Female-First/Last Author (1956/69)	-0.131***	-0.124***	-0.098***	-0.128***	-0.098***	-0.129***	-0.102***
	(0.019)	(0.017)	(0.015)	(0.017)	(0.018)	(0.016)	(0.017)
		[0.025]	[0.019]	[0.023]	[0.022]	[0.024]	[0.022]
Observations	255,768	255,768	255,768	255,768	255,768	255,768	255,768
R^2	0.009	0.470	0.469	0.499	0.498	0.513	0.511
Sample 3: Prestigious Universities, Sciences, 1900-2000							
Female-First/Last Author (1900/14)	-0.172	-0.160	-0.087	-0.158**	-0.064*	-0.245***	-0.142***
	(0.137)	(0.117)	(0.073)	(0.074)	(0.035)	(0.056)	(0.020)
		[0.123]	[0.07]	[0.101]	[0.055]	[0.093]	[0.05]
Female-First/Last Author (1925/38)	-0.110**	-0.104**	-0.093***	-0.064	-0.062*	-0.075	-0.074**
	(0.055)	(0.044)	(0.027)	(0.050)	(0.035)	(0.047)	(0.031)
		[0.06]	[0.031]	[0.065]	[0.036]	[0.061]	[0.034]
Female-First/Last Author (1956/69)	-0.162***	-0.151***	-0.136***	-0.143***	-0.126***	-0.144***	-0.126***
	(0.029)	(0.026)	(0.022)	(0.026)	(0.024)	(0.027)	(0.025)
		[0.033]	[0.024]	[0.03]	[0.023]	[0.03]	[0.024]
Female-First/Last Author (2000)	-0.079***	-0.051***	-0.046***	-0.053***	-0.047***	-0.055***	-0.047***
	(0.010)	(0.007)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)
		[0.009]	[0.005]	[0.009]	[0.005]	[0.009]	[0.005]
Observations	611,513	611,513	611,513	611,513	611,513	611,513	611,513
R^2	0.016	0.375	0.391	0.402	0.416	0.411	0.425
Predicted Citation Control		Yes		Yes		Yes	
Predicted Citation Control (All)			Yes		Yes		Yes
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes	Yes		
Cohort×University FE				Yes	Yes		
Cohort×Discipline×University FE						Yes	Yes

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

Importantly, papers by female scientists were not under-cited because women worked on topics with less citation potential but rather because of other biases in the citation market. Estimated citation gaps hardly change if we control for our novel measure of predicted citations. This holds whether we predict citations using papers by male scientists (column 2), or by all scientists (column 3).

Citations gaps are comparable for papers published by female scientists from prestigious universities (sample 3, see also Appendix Figure D.1). For the 2000 cohort, the estimated citation gaps are around 0.05 s.d., indicating that citation gaps still persisted at the end of the 20th century, but had significantly shrunk. Moreover, citation gaps are similar if we compare papers published by scientists in the same cohort and university or even the same cohort, university, and discipline (e.g., Harvard biology in 2000).

Robustness

The results are similar if we use out-of-sample predictions of citations using a cross-fitting procedure (Appendix Table D.1, columns 2-3). For this procedure, we divide the data into k slices. We then train the model using $k-1$ slices. Once the model is trained, we predict citations for the left-out slice. We repeat the procedure until we obtain predictions for all k slices. The results are also robust to using alternative functional forms for the predicted citation control, i.e. controlling linearly for predicted citations (column 4) or with 1,000 indicators for the permilles of the predicted citation distribution (column 5). The results also remain unchanged if we do not winsorize citations (column 6). Further, we show that results are robust when using citation counts, instead of standardized citations, as the dependent variable (column 7). Lastly, we show that our findings also remain unchanged if we use a double machine learning approach (Chernozhukov et al. 2018), which directly controls for the words from the titles in a version of regression (14) (column 8).[48]

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

Alternative Explanations

Apart from the fact that women may have worked on topics with different citation potential, there are at least two additional reasons that may explain why papers by female scientists received fewer citations. First, women may have fewer opportunities to write papers with coauthors (production effect). This may translate into fewer citations because coauthored papers, on average, receive more citations (e.g., Wuchty et al. 2007). Second, women may publish their papers in lower-ranked journals because of biased editors or referees (publication effect). Such an effect has been shown for economics papers (Card et al. 2022). 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). Controlling for the number of authors does not affect gender gaps in citations (Table 6, columns 1-3). Next, we explore the publication effect by including a full set of journal fixed effects (columns 4-6). The inclusion of these has little impact on the magnitude of the estimates. Similar to universities, disciplines, or titles, the number of coauthors and the journal are potentially endogenous. Therefore, these results should be interpreted as a decomposition of the citation gaps.

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

Table 6: Gender Gaps in Citations: Accounting for Number of Authors and Journals

	Author Nr. FE			Journal FE		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Standardized Citations					
Sample 2: All Universities, Sciences, 1900-1969						
Female-First/Last Author (1900/14)	-0.142*	-0.156**	-0.186***	-0.179**	-0.206***	-0.246***
	(0.080)	(0.071)	(0.067)	(0.070)	(0.058)	(0.058)
	[0.131]	[0.123]	[0.126]	[0.129]	[0.117]	[0.118]
Female-First/Last Author (1925/38)	-0.128***	-0.098**	-0.093**	-0.131***	-0.095**	-0.100**
	(0.035)	(0.041)	(0.039)	(0.038)	(0.042)	(0.039)
	[0.051]	[0.063]	[0.063]	[0.052]	[0.06]	[0.06]
Female-First/Last Author (1956/69)	-0.118***	-0.122***	-0.121***	-0.103***	-0.122***	-0.121***
	(0.017)	(0.017)	(0.016)	(0.019)	(0.016)	(0.015)
	[0.025]	[0.023]	[0.024]	[0.026]	[0.022]	[0.023]
Observations	255,768	255,768	255,768	255,768	255,768	255,768
R^2	0.471	0.500	0.514	0.493	0.521	0.533
Sample 3: Prestigious Universities, Sciences, 1900-2000						
Female-First/Last Author (1900/14)	-0.158	-0.150**	-0.237***	-0.171	-0.184***	-0.312***
	(0.111)	(0.074)	(0.054)	(0.127)	(0.063)	(0.032)
	[0.119]	[0.101]	[0.093]	[0.13]	[0.093]	[0.081]
Female-First/Last Author (1925/38)	-0.105**	-0.064	-0.075	-0.093**	-0.053	-0.073
	(0.044)	(0.053)	(0.050)	(0.043)	(0.050)	(0.047)
	[0.061]	[0.066]	[0.063]	[0.06]	[0.063]	[0.06]
Female-First/Last Author (1956/69)	-0.145***	-0.136***	-0.137***	-0.134***	-0.142***	-0.146***
	(0.026)	(0.027)	(0.027)	(0.028)	(0.026)	(0.025)
	[0.033]	[0.03]	[0.03]	[0.034]	[0.028]	[0.027]
Female-First/Last Author (2000)	-0.037***	-0.039***	-0.040***	-0.039***	-0.042***	-0.044***
	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
	[0.009]	[0.009]	[0.009]	[0.008]	[0.008]	[0.008]
Observations	611,513	611,513	611,513	611,513	611,513	611,513
R^2	0.379	0.405	0.414	0.436	0.461	0.469
Predicted Citations Control	Yes	Yes	Yes	Yes	Yes	Yes
Cohort×Discipline×Country FE	Yes	Yes		Yes	Yes	
Cohort×University FE		Yes			Yes	
Cohort×Discipline×University FE			Yes			Yes
Nr. Authors FE	Yes	Yes	Yes			
Journal FE				Yes	Yes	Yes

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

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

5 Gender Gaps in Promotions

In this 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 data in cohort $t-1$.[49] We then analyze whether they get promoted to full professor by cohort t (see Appendix A.1.1. for details on the coding of promotions). Promotions to full professor are particularly important because, in all countries, full professors have unique privileges and high job security and salaries. We estimate the following regression:

$$
\begin{aligned}
\text{Promotion Full Prof}_{it} &= \pi_1 + \pi_2 \text{Female}_i \times 1[t(i) = 1914] + \pi_3 \text{Female}_i \times 1[t(i) = 1925/38] \\
&+ \pi_4 \text{Female}_i \times 1[t(i) = 1956/69] + \pi_5 \text{Female}_i \times 1[t(i) = 2000] \\
&+ \text{FE}(i, t-1) + \upsilon_{it}.
\end{aligned} \tag{15}
$$

The dependent variable Promotion Full Prof_{it} is an indicator that equals 1 if academic i, who entered the data in cohort $t-1$, was promoted to full professor by cohort t.[50] The main explanatory variables are the interactions of the indicator Female_i with indicators for the four different time periods. The regressions include fixed effects as defined in (2), evaluated in $t-1$. The fixed effects control, for example, for the fact that in certain time periods, disciplines, and universities, there may have been more full professor openings.

In all universities and disciplines (sample 1), women who started their careers in the 1900 cohort were, on average, 13 percentage points less likely than men to be promoted to full professor by 1914 (Table 7, sample 1, column 1).[51] Because the probability of promotion to full professor was around 16% in 1914, women were about 79% less likely to be promoted. Women in the 1925 and 1938 cohorts and those in the 1956 and 1969 cohorts were around 14 and 12 percentage points (or between 87% and 76%) less likely than men to be promoted to full professor by the next cohort. The large gender gap in promotions to full professor is robust to the inclusion of more stringent fixed effects. We estimate similar gender gaps in promotions if we compare women and men who started their careers in the same cohort and university or even in the same cohort, university, and discipline (Table 7, sample 1, column 3).

In the scientist sample (mathematics, physics, chemistry, biochemistry, and biology) of all universities (sample 2), women who started their careers in 1900 were around 14 percentage points less likely to be promoted by 1914. Women in the 1925-1938 cohorts and in the 1956-1969 cohorts were 11 and 18 percentage points less likely than men

[49]This restriction results in a smaller sample because academics who enter the data as full professors are not included in the analysis. Furthermore, all academics who enter the data in the last cohort (independently of their rank) are also not included in the analysis.

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

[51]Gender gaps in promotions are driven by two factors: first, women are less likely to apply for promotion and, second, conditional on applying they are less likely to be promoted.

Table 7: Gender Gaps in Promotions

	(1)	(2)	(3)	(4)
Dependent Variable:	Indicator of Promotion to Full Professor			
Sample 1: All Universities, all disciplines, 1900-1969				
Female (1914)	-0.127***	-0.126**	-0.088*	
	(0.037)	(0.049)	(0.046)	
Female (1925/38)	-0.138***	-0.149***	-0.142***	
	(0.015)	(0.016)	(0.017)	
Female (1956/69)	-0.122***	-0.103***	-0.103***	
	(0.014)	(0.010)	(0.011)	
Observations	102,611	102,611	102,611	
R-squared	0.114	0.223	0.388	
Sample 2: All Universities, sciences, 1900-1969				
Female (1914)	-0.135	0.342**	0.266	0.268
	(0.128)	(0.171)	(0.169)	(0.171)
Female (1925/38)	-0.114***	-0.168***	-0.151***	-0.147***
	(0.038)	(0.044)	(0.042)	(0.042)
Female (1956/69)	-0.184***	-0.153***	-0.160***	-0.153***
	(0.015)	(0.021)	(0.020)	(0.020)
Std. Publications				0.046***
				(0.006)
Std. Citations				0.010
				(0.006)
Observations	16,573	16,573	16,573	16,573
R-squared	0.103	0.300	0.457	0.465
Sample 3: Prestigious Universities, sciences, 1900-2000				
Female (1914)	-0.193	0.810***	0.600***	0.604***
	(0.175)	(0.230)	(0.227)	(0.229)
Female (1925/38)	-0.169***	-0.218***	-0.186***	-0.176***
	(0.042)	(0.054)	(0.059)	(0.059)
Female (1956/69)	-0.215***	-0.234***	-0.233***	-0.223***
	(0.021)	(0.030)	(0.021)	(0.023)
Female (2000)	-0.095***	-0.071***	-0.067***	-0.058***
	(0.028)	(0.017)	(0.016)	(0.015)
Std. Publications				0.030***
				(0.007)
Std. Citations				0.018***
				(0.005)
Observations	12,580	12,580	12,580	12,580
R-squared	0.154	0.290	0.419	0.425
Cohort×Discipline×Country FE	Yes	Yes		
Cohort×University FE		Yes		
Cohort×Discipline×University FE			Yes	Yes

Notes: The Table shows gender gaps in the probability of promotion to full professor. Results are estimated at the academic-level. Sample 1 includes academics in all disciplines and all universities until 1969. Sample 2 includes scientists (mathematics, physics, chemistry, biochemistry, and biology) in all universities until 1969. Sample 3 includes scientists in prestigious universities until 2000. The dependent variable is an indicator that equals 1 if an academic who entered the dataset in cohort $t-1$ at a lower rank than full professor was promoted to full professor by cohort t, or 0 otherwise. The main explanatory variable is an indicator that equals 1 if the academic is a woman, interacted with the relevant cohort(s). The regressions also control for different sets of fixed effects (see definition (2) for details) evaluated in $t-1$. Standard errors are clustered at the discipline-country level, with 1,606 clusters in sample 1, 285 in sample 2, and 161 in sample 3. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

to be promoted to full professor (Table 7, sample 2, column 1). For the cohorts after 1925, the promotion gaps are similar if we condition on more stringent fixed effects. For the 1900 cohort (i.e., those who could have been promoted by 1914), the effects turn positive if we include more stringent fixed effects. However, given the extremely low female representation in the sciences in 1900, only 16 women in all universities of the world combined could have been promoted to full professor. Comparisons in promotions within the same cohort and university or cohort, university, and discipline are thus based on only a handful of women.

Results are also similar if we estimate promotion gaps for scientists in prestigious universities (sample 3). In this sample, we can extend the time horizon and find that promotion gaps have declined to about 6 to 10 percentage points by 2000.

In the second and third parts of the paper, we have shown that women published fewer papers and received fewer citations throughout the 20th century. To explore whether gender gaps in publications and citations affect gender gaps in promotions, we add controls for scientists' publication and citation records.[52] A one standard deviation better publication record increased the probability of promotion to full professor by 4.6 (sample 2) or 3 (sample 3) percentage points. A one standard deviation better citation record did not significantly increase the probability of promotion in sample 2, but it increased it by 1.8 percentage points in sample 3. The small effect of citations on promotions in sample 2 is likely driven by the fact that only since the 1960s it has been possible to systematically measure citations (Hager et al., 2023).

Strikingly, controlling for the publication and citation records hardly affects the estimated gender gaps in the promotion to full professor.[53] The unexplained part of the promotion gap is larger than the effect of a three to seven s.d. worse publication record. This is remarkable because the true quality of women, conditional on the number of publications and citations should be, if anything, higher in the presence of discrimination and other biases in the publication market.

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

6 Conclusion

Leveraging new hand-collected worldwide data, this paper sheds light on the evolution of gender gaps in academia over the 20th century. From our analysis, four results stand out.

First, the share of women in academia very low throughout the 20th century. It was around 1% in 1900 and increased by about 1.5% per decade until the 1960s. Female shares were especially low in STEM disciplines and in prestigious universities. From the 1970s until the year 2000, female shares increased substantially. However, in the year

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

[53] In unreported results, we control more flexibly for publications and citations by including indicators for various percentiles of the distributions of publications and citations or by including polynomials of the standardized numbers of publications and citations. The results are similar to those reported in Table 7.

2000, women remained underrepresented. Second, women published fewer papers than men. We outline a model that links gender gaps in hiring and gender gaps in publishing. We estimate the model using cross-country and over-time variation in the shares of women in academia. We find a U-shaped relationship between gender gaps in hiring and gender gaps in publications (the "gender U"), demonstrating that these gaps are inherently linked. With rising female shares, the relative importance of positive selection of women was offset by increased publishing opportunities for women. Third, papers by female authors received fewer citations. We develop a novel machine learning approach that shows that citation gaps do not arise because women work on topics that are generally less cited. Fourth, female academics were less likely to be promoted to full professor. Strikingly, the promotion gap does not stem from gender gaps in publications and citations.

This article provides the first comprehensive analysis of gender gaps in a high-skilled profession at a global scale and covering the entire 20th century. Overall, the results document pervasive unequal opportunities for women in academia throughout the 20th century. The broad perspective taken in this article provides the foundation for a more systematic understanding of gender gaps in academia and highlights fruitful directions for future research and science policy.

References

Albanesi, S., C. Olivetti, and M. J. Prados (2015). Gender and Dynamic Agency: Theory and Evidence on the Compensation of Top Executives. In *Gender in the Labor Market*, pp. 1–59. Emerald Group Publishing Limited.

Altonji, J. G. and R. M. Blank (1999). Race and Gender in the Labor Market. *Handbook of Labor Economics 3*, 3143–3259.

Amemiya, T. (1984). Tobit Models: A Survey. *Journal of Econometrics 24*(1-2), 3–61.

Ashraf, N., O. Bandiera, V. Minni, and V. Quintas-Martínez (2022). Gender Roles and the Misallocation of Labour Across Countries. *mimeo LSE*.

Azmat, G. and R. Ferrer (2017). Gender Gaps in Performance: Evidence from Young Lawyers. *Journal of Political Economy 125*(5), 1306–1355.

Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017). Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. *American Economic Review 107*(3), 714–747.

Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017). Does the Gender Composition of Scientific Committees Matter? *American Economic Review 107*(4), 1207–1238.

Barocas, S. and A. D. Selbst (2016). Big Data's Disparate Impact. *Calif. L. Rev. 104*, 671.

Becker, S. O., V. Lindenthal, S. Mukand, and F. Waldinger (2023). Scholars at Risk: Professional Networks and High-Skilled Emigration from Nazi Germany. *American Economic Journal: Applied Economics (forthcoming)*.

Beidas, R. S., P. A. Hannon, A. S. James, and K. M. Emmons (2022). Advancing Gender Equity in the Academy. *Science Advances 8*(13), eabq0430.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *The Quarterly Journal of Economics 134*(2), 647–713.

Bertrand, M. (2011). New Perspectives on Gender. In *Handbook of Labor Economics*, Volume 4, pp. 1543–1590. Elsevier.

Bertrand, M., S. E. Black, S. Jensen, and A. 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, M. and E. Duflo (2017). Field Experiments on Discrimination. *Handbook of Economic Field Experiments 1*, 309–393.

Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics 2*(3), 228–255.

Bertrand, M. and K. F. Hallock (2001). The Gender Gap in Top Corporate Jobs. *ILR Review 55*(1), 3–21.

Blau, F. D. and L. M. Kahn (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature 55*(3), 789–865.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2022). Gender Differences in Peer Recognition by Economists. *Econometrica 90*(5), 1937–1971.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2023). Gender Gaps at the Academies. *Proceedings of the National Academy of Sciences 120*(4), e2212421120.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal 21*(1), C1–C68.

Ciardi, M. and M. Focaccia (2011). Maria Bakunin (1973-1969). In *European Women in Chemistry*. Wiley Online Library.

Curie, E. (1938). *Madame Curie*. Paris, France: Gallimard.

De la Croix, D., F. Docquier, A. Fabre, and R. Stelter (2023). The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000–1800). *Journal of the European Economic Association*, jvad061.

Gayle, G.-L., L. Golan, and R. A. Miller (2012). Gender Differences in Executive Compensation and Job Mobility. *Journal of Labor Economics 30*(4), 829–872.

Goldin, C. (1989). Life-Cycle Labor-Force Participation of Married Women: Historical Evidence and Implications. *Journal of Labor Economics 7*(1), 20–47.

Goldin, C. (1990). *Understanding the Gender Gap: An Economic History of American Women*. Oxford, GB: Oxford University Press.

Goldin, C. and L. F. Katz (2016). A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation. *Journal of Labor Economics 34*(3), 705–746.

Hager, S., C. Schwarz, and F. Waldinger (2023). Measuring Science: Performance Metrics and the Allocation of Talent. *CEPR Discussion Paper* (18248).

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161.

Hengel, E. (2020). Publishing while Female (Summary).

Hill, R. and C. Stein (2021). Race to the Bottom: Competition and Quality in Science. *Northwestern University and UC Berkeley*.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The Allocation of Talent and US Economic Growth. *Econometrica 87*(5), 1439–1474.

Huang, J., A. J. Gates, R. Sinatra, and A.-L. Barabási (2020). Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines. *Proceedings of the National Academy of Sciences 117*(9), 4609–4616.

Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science. *The Quarterly Journal of Economics 133*(2), 927–991.

Jones, C. I. (1995). R & D-Based Models of Economic Growth. *Journal of Political Economy 103*(4), 759–784.

Killingsworth, M. R. and J. J. Heckman (1986). Female Labor Supply: A Survey. *Handbook of Labor Economics 1*, 103–204.

Koffi, M. (2021). Innovative Ideas and Gender Inequality. Technical report, Working Paper Series.

Kricheldorf, H. R. (2014). *Erkenntnisse und Irrtümer in Medizin und Naturwissenschaften*. Berlin, Germany: Springer-Verlag.

McGrayne, S. B. (1998). *Nobel Prize Women in Science: their Lives, Struggles, and Momentous Discoveries*. New York, N.Y.: Birch Lane Press, Carol Publishing Company.

Mitchell, B. R. (1982). *International Historical Statistics: Africa and Asia*. London, GB: The Macmillan Press LTD.

Mitchell, B. R. (1983). *International Historical Statistics: The Americas and Australasia*. London, GB: The Macmillan Press LTD.

Mitchell, B. R. (1993). *International Historical Statistics: Europe 1750-1988*. New York, N.Y.: Stockton Press.

Moser, P. and S. Kim (2021). Women in Science. Lessons from the Baby Boom. Technical report, National Bureau of Economic Research.

Mulligan, C. B. and Y. Rubinstein (2008). Selection, Investment, and Women's Relative Wages over Time. *The Quarterly Journal of Economics 123*(3), 1061–1110.

Nature (1930). Index Generalis: Annuaire Général des Universités, Grandes Écoles, Académies, Archives, Bibliothèques, Institute Scientifiques, Jardins Botaniques et Zoologiques, Musées, Observatoires, Stés Savantes. *Nature 126*, 126–127.

OECD (2021). Entrepreneurship - Inventors. `https://data.oecd.org/entrepreneur/inventors.htm`.

Olivetti, C. and B. Petrongolo (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics 8*, 405–434.

Romanowicz, B. (2019). Equality of Opportunity in the European Research Council Competitions. In *EGU General Assembly Conference Abstracts*, Volume 10306.

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

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

Ross, M., B. Glennon, R. Murciano-Goroff, E. Berkes, B. Weinberg, and J. Lane (2022). Women are Credited Less in Science than are Men. *Nature 608*(7921), 135–145.

Rossiter, M. W. (1982). *Women Scientists in America: Struggles and Strategies to 1940*, Volume 1. Baltimore, MD: JHU Press.

Roussille, N. (2021). The Central Role of the Ask Gap in Gender Pay Inequality. *University of California, Berkeley.*

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers 3*(2), 135–146.

Santamaría, L. and H. Mihaljević (2018). Comparison and Benchmark of Name-to-Gender Inference Services. *PeerJ Computer Science 4*, e156.

Sarsons, H. (2017a). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. *mimeo UBC*, 141–145.

Sarsons, H. (2017b). Recognition for Group Work: Gender Differences in Academia. *American Economic Review 107*(5), 141–145.

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2021). Gender Differences in Recognition for Group Work. *Journal of Political Economy 129*(1), 101–147.

Schwarz, C. (2023). Estimating Text Regressions Using txtreg_train. *The Stata Journal 23*(3), 799–812.

Shanghai Ranking (2020). Academic Ranking of World Universities 2020. `https://www.shanghairanking.com/rankings/arwu/2020`.

Shepley, C. F. (2008). *Movers and Shakers, Scalawags and Suffragettes: Tales from Bellefontaine Cemetery.* St.Louis, MO: Missouri History Museum.

Sinha, A., Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. Hsu, and K. Wang (2015). An Overview of Microsoft Academic Service (MAS) and Applications. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 243–246.

Truffa, F. and A. Wong (2022). Undergraduate Gender Diversity and Direction of Scientific Research. *Work. Pap., Stanford Univ., Stanford, CA.*

Vaughn, G. F. (2004). Katharine Coman: America's First Woman Institutional Economist and a Champion of Education for Citizenship. *Journal of Economic Issues 38*(4), 989–1002.

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

WHED (2024). World higher education database (whed) portal. Accessed on 2024-06-17.

Wigner, E. P. (1972). Obituaries - Maria Goeppert Mayer. *Physics Today 25*(5), 77.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science 316*(5827), 1036–1039.

Online Appendix

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

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

A Further Details on Data

A.1. Enhancements of Faculty Roster Data

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

Minerva and the university websites report academic ranks for most academics. The ranks are reported either in the original language (e.g., maître de conférence) or are translated into English or German. Overall, the sources report almost 4,000 different combinations of countries and ranks. We recode them at the country level, because certain labels of ranks do not necessarily describe the same academic rank across countries. 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 list a academics who hold different academic ranks under only one heading (e.g., a joint heading of "associate and assistant professors" instead of listing "associate professors" and "assistant professors" under separate headings). In these cases, we assign the highest listed rank to each academic.

In many academic systems, e.g., in Germany and Switzerland, young researchers climb the academic ladder by substituting 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 and 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 and research position).

Promotion to Full Professor

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

A.1.2. Additional Information on the Coding of Disciplines

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

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

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

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

We identify academics with multiple appointments within a city by hand-checking all academics with duplicate surnames within a city. We then determine whether two entries refer to the same academic based on the first name, specialization, academic rank, and title. If an academic holds two appointments, we harmonize the first name and collapse the two entries into a single observation. The resulting observation then contains the information on all appointments and specializations of an academic within a city.[54]

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

A.1.4. Linking Academics Across Cohorts

As described in the main text, we link academics across cohorts. 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 variation in first names occurs 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

We first link academics who remain in the same department between cohorts t and $t+1$. In a first step, we obtain potential links by merging academics from discipline d, country c, and university u, in cohort t to academics from the same discipline d, same country c, and same university u in cohort $t+1$ based on the academic's surname and 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 classify 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 academics in 1925 and 1938:

Table A.1: Examples within Department Merge

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

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

Linking Academics across Departments in the same Country

Second, we link academics who remain in the same country but change departments between two cohorts. In the first step, we obtain potential links by merging academics from discipline d, country c, cohort t to academics from the same discipline d, same country c but cohort $t+1$ based on the academic's surname, the first initial. 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.[55] We then process the potential links as follows:

1. If the entire information on the first name is identical in both cohorts, we classify 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 correct, the research assistants use the following rules:

[55] A small number of universities change 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 from 1925 onward. Hence, the within-country link for the University of Strasbourg considers academics who moved from or to other German universities between 1900 and 1914, as well as between 1914 and 1925. It also considers academics who moved from or to other French universities between 1914 and 1925 and all following cohorts. The moves from Germany to Strasbourg or from Strasbourg to France between 1914 and 1925 (when the university changed country) are considered in the cross-country link that we describe below.

(a) If there are only minor spelling differences in the first name, the research assistant classifies the potential link as correct (see lines 1 and 2 in 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 correct (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 u' before the year corresponding to cohort $t+1$, the potential link is classified as correct. E.g., K(arl) Röder (see lines 5 and 6 in Table A.2) could be found online and his Wikipedia entry states that:

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

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

Table A.2: Within Country Merge

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

Linking Academics across Countries

Third, we link academics who move across countries. In the first step, we obtain potential links by merging academics with the same surname, first initial, and discipline d in cohort t to academics with the same surname, first initial, and discipline d, in cohort $t+1$ who are listed in two different countries. To rule out false positives, all potential links are confirmed by extensive manual online searches. 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 correct.

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

For most academics, we infer their gender on the basis of their first name and their country.[56] 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. After linking the observations, we adjust the first name in 1925 to Henry C and can therefore code the gender of Cowles in 1925.

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

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

A.2. Additional Information on Coding Gender

A.2.1. Example Google Picture Search

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

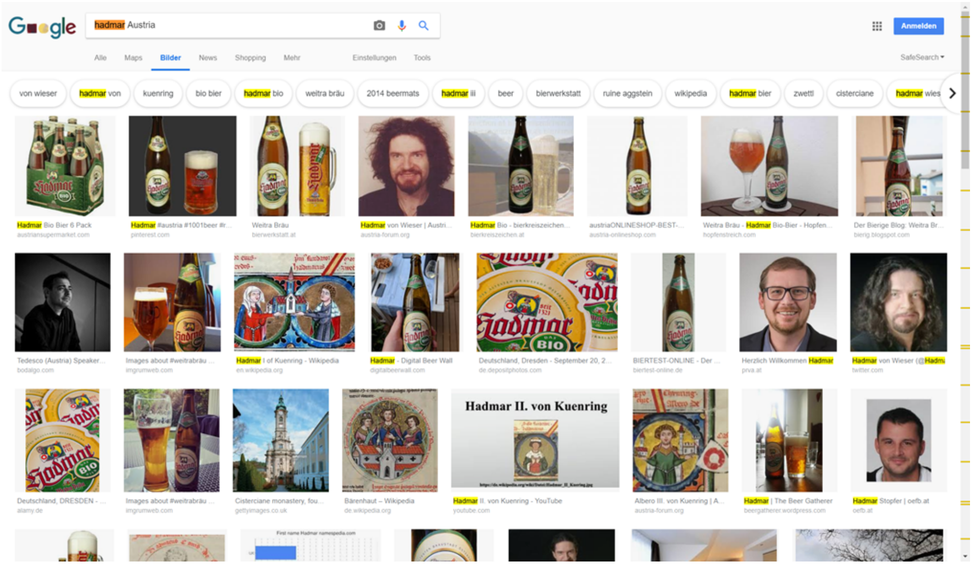
A.2.2. Hand-Checking Gender Coding

As described in the main text, in the last step of assigning gender, we hand-check individual academics who appear misclassified. Such misclassifications occur mostly because the predominant gender of a first name - country combination changes over time.[57] 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 academic 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 (1861-1936), Foreign Member of the Society, was born in Angers on 12

[56] 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).

[57] 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.

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. For example, could be positive if the most talented women were both more likely to get academic positions and to publish well once hired.ombination (see section A.3. for details).

May 1861. *He* studied at Montpellier...” (see here for details). This allows us to identify him as male.

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

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.”[58] 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).

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

5. We further extract the first of the initials from the list of initials obtained in the previous step.

A.3.2. Preparing Addresses in 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 instances, an alternative database, Microsoft Academic Graph (MAG), contains the relevant address information. We therefore enrich the affiliations as reported by the Web of Science with information from MAG.[59] 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 changed their name (e.g., St. Petersburg became Leningrad), and a number of cities changed countries (e.g., Strasbourg was German in 1900 and 1914, and then became French). Furthermore, cities may be spelled in different languages in the faculty rosters and 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 it is spelled with either Italian ("Roma"), English ("Rome"), or German

[59] 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 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.

("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 the faculty rosters using Google Maps. The first step relies on the Google Maps API.

Step 1, part (i). We submit all city-country pairs (e.g., "London, United Kingdom") that appear in the Web of Science to the API.[60] 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 rosters 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 name of a city or a country has changed over time (e.g., the name of Preßburg changed into Bratislava) or because of typos in the Web of Science. We can identify such cases because the location type flag is "APPROXIMATE" instead of "CITY". We improve the geocoding for these cases using the following 3-step procedure:

1. We structure the address before re-submitting it to the API, (e.g., "'city' : Preßburg, 'country' : Hungary").[61]
2. For the cases without a result in step one, we re-format the address-city string (e.g., "Preßburg,+Hungary") and then re-submit it to the API.
3. For the cases that do not return a result in steps one and two, we re-submit the complete address (not just the city and country) from the Web of Science (e.g., "Loyola Univ Clinics, Mercy Hosp, Chicago, IL USA") to the API.

Step 2. In some cases, the procedure above does not guarantee that the correct city and country 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.

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

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

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 finds the address, we extract the city and country from the website and use them as inputs for the Google Maps API (i.e. Step 1, part i). We further process the output from the Google Maps web interface with 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 fields (e.g., physics or general science) based on the journal they are published in, as opposed to assigning fields at the paper level. For 59% of the papers, this establishes a unique assignment to one of the disciplines in the faculty rosters (e.g., the journal Acta Mathematica is uniquely assigned to mathematics). The remaining 41% of papers are published in journals that the Web of Science either assigns to multiple disciplines (e.g., the journal Biometrika is assigned to mathematics as well as biology) or to general science (e.g., Nature).[62] Matching the general science 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 logit 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 using the words (unigrams), word pairs (bigrams), and word triplets (trigrams) from the titles of the 15,078,761 papers that the Web of Science assigns to unique disciplines (e.g., the papers published in Acta Mathematica). In preparation for the training of the classifier, we remove very common words (stopwords) from the titles, as these contain little information. Next, we reduce words to their morphological roots using a stemmer. Afterwards, we transform the titles of each paper into a document 1,2,3-gram matrix $\mathbf{X}$ of dimension $D \times V$, where D is the number of papers in our data and V is the total number of unique unigrams, bigrams, and trigrams in all titles:

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

[62] 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 $\text{tf-idf}(w_{d,v}) = (1 + \log(w_{d,v})) \times \left(\log\left(\frac{1+D}{1+d_v}\right) + 1\right)$, where d_v is the number of documents n-gram v appears in at least once. This reweighting reduces the weights of n-grams that appear in many titles of papers (e.g., method).

The multinomial logit classifier then learns to predict disciplines based on the 1,2,3–gram matrix $\mathbf{X}$, where the dependent variable y_d is the discipline of the paper. To avoid overfitting, we include L2 regularization in the classifier. As is standard in the machine learning literature, we choose the optimal regularization strength using 10-fold cross-validation, evaluated on the basis of the F1-score.[63] The final classifier achieves a within-sample F1-score of 0.99 and an out-of-sample F1-score of 0.81. As some biochemistry papers are published in chemistry journals, we would expect an F1 score of less than one. Using our classifier, we predict a unique discipline for the 10,508,299 papers which the Web of Science originally assigns to multiple disciplines (on the basis of the journal).

A.4. Merging Academics to Web of Science

We match papers from the Web of Science to the faculty rosters using a nine-step procedure. As mentioned in the main text, we match papers from the Web of Science within a ± 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. 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 incorrect matches: Denote the string of initials of a scientist in the faculty rosters by s and the string of initials of the scientist in the Web of Science by p:

   (a) If the number of initials in s and p is identical ($|s| = |p|$), but the initials differ ($s \neq p$) we exclude the match. For example, a match of a scientist listed in the faculty rosters with initials "A.A." will not be merged to a paper published by

[63] 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.

someone with initials "A.B." (Note: as described in step 3, we only consider matches where the full surname, subject, country, and city match.)

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

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

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

If one of the authors of a paper is matched to a scientist in an earlier (and thus more restrictive) step, this particular author will no longer be considered in any following step. We account for the fact that some papers are merged to multiple scientists by weighting the papers by the total number of matches. For the time period covered by our paper, the Web of Science rarely provides a unique assignment of the addresses reported on a paper to its coauthors: e.g., if a paper has two coauthors, and they are affiliated with different institutions, usually the Web of Science does not specify which coauthor is affiliated with which institution. We, therefore, assign each address reported in a paper to all coauthors of the paper. If more than one address is associated with a paper, we perform a many-to-many merge of addresses to coauthors. As we show in Table 3, the results are robust to only considering scientists who have a unique surname, first initial, and discipline in all universities of the world.

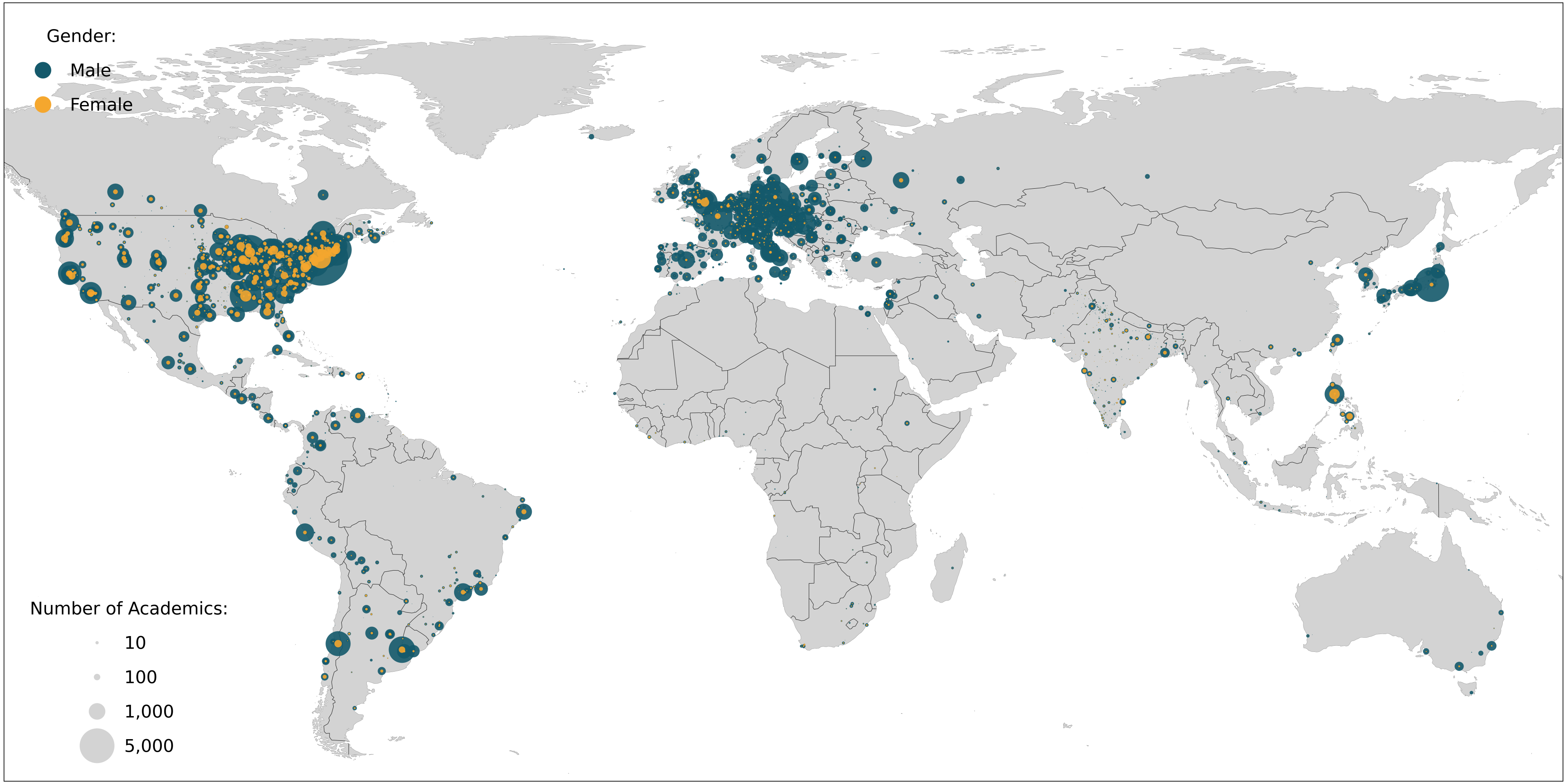
A.5. Descriptive Information: Minerva Data

Table A.3: Universities in Minerva and WHED

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

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

Figure A.2: Academics in Minerva 1900-1969



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

A.6. Benchmarking the Minerva Data

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 in two ways. First, we show that the number of universities covered in Minerva is similar to the number of universities included in the World Higher Education Database (WHED (2024), see Appendix Table A.3). The comparison indicates that Minerva captures a very large fraction of worldwide universities. For most countries, Minerva covers more universities than are listed in the World Higher Education Database. Because the Soviet Union stopped reporting academics for Minerva from the 1938 cohort onwards, the coverage is worse for Russia and other countries that were part of the Soviet Union. The coverage is also lower in China, which only established a modern university sector during the course of the 20th century.

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

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

Rossiter (1982), 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.

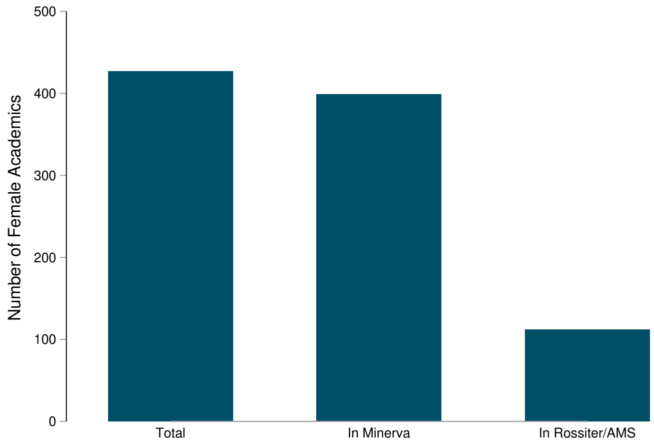
For the benchmarking exercise, we extract all female scientists who are at least assistant professors that are listed in these twenty universities in Minerva 1938. We then cross-check all names and identify women listed in both sources. Both sources combined list a total of 427 different female academics, which we take as the best available information for the total number of women in these twenty universities in 1938 (first bar, Figure A.3. Of these, 399 (93%) are listed in Minerva (second bar).[64] In contrast, Rossiter, on the basis of the American Men of Science, only lists 112 (26%) of them (third bar). This indicates that Minerva 1938 has a much more comprehensive coverage of academics in the top twenty U.S. universities for 1938 than the American Men of Science.

A.6.2. Benchmarking Against German University Catalog Data

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

[64]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.

Figure A.3: Benchmarking Minerva Against Rossiter (1982) / American Men of Science (1938)



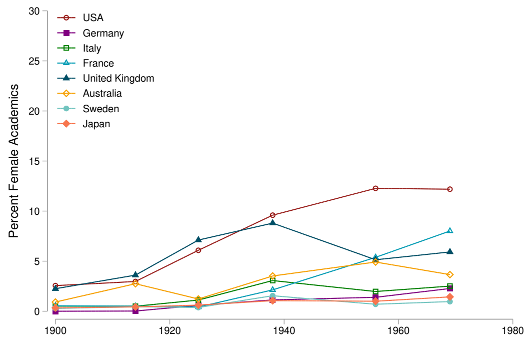
Notes: The Figure shows the number of female scientists in twenty major U.S. universities for the year 1938 and how they are covered by different sources.

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

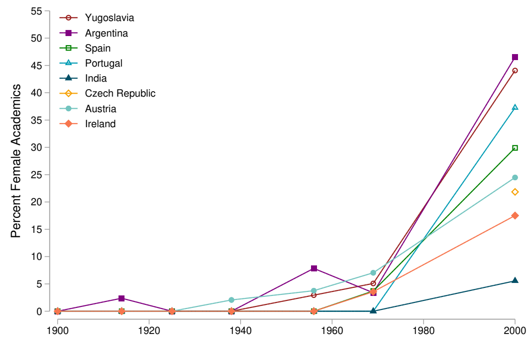
B Further Results: Hiring Gaps

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

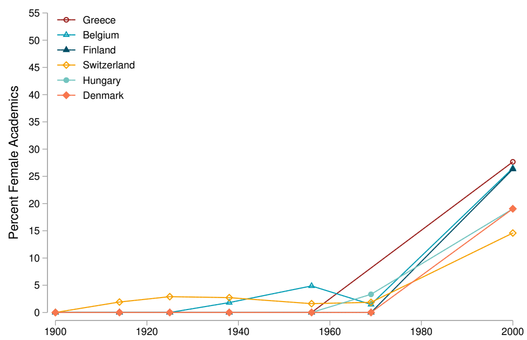
(a) Excluding Women's Colleges



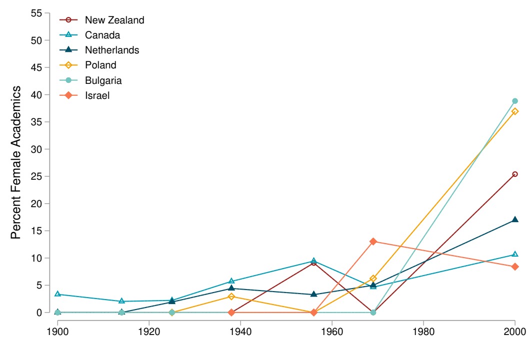
(b) Additional Countries



(c) Additional Countries

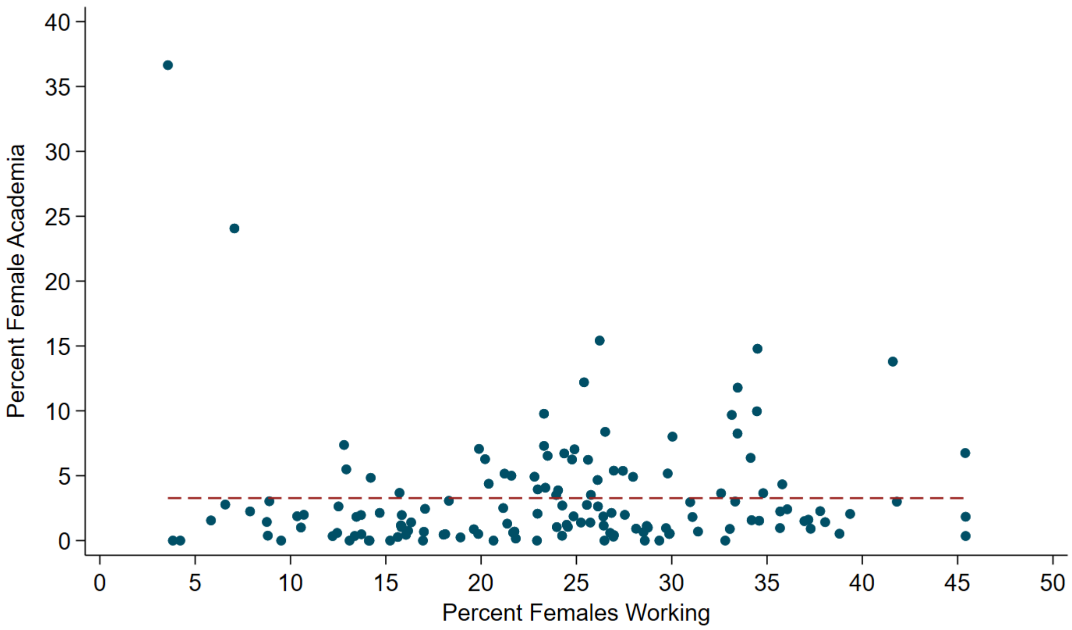


(d) Additional Countries



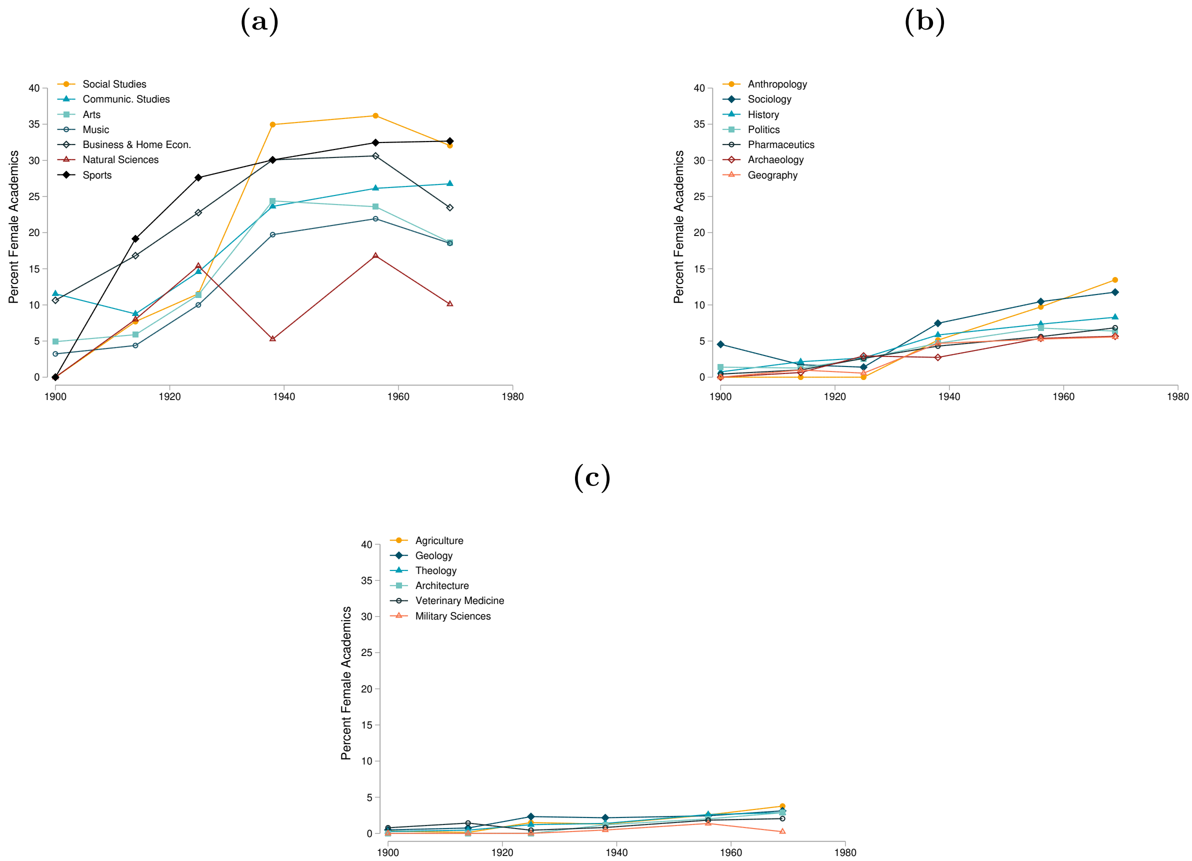
Notes: The Figure shows the percentage of female academics by country over time. Panel (a) plots the percentage of female academics across all universities and disciplines until 1969 (sample 1), excluding women's colleges. Panels (b)-(d) plot the percentage of female scientists in prestigious universities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

Figure B.2: Correlation of Percent Female Academics with General Labor Force Participation of Women



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

Figure B.3: Percent of Female Academics, Additional Disciplines



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 Home Economics into one line. The data were collected by the authors from various volumes of Minerva and department websites, see section 1 for details.

C A More General Roy Model

In the more general version of the 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 is a possible gender bias in hiring, and ϵ_{0i}^{g} is the unobserved component of these latent valuations. As a result, academic position i is filled with a woman if:

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

with $\epsilon_{0i} \equiv \epsilon_{0i}^{W} - \epsilon_{0i}^{M}$, $X_i \equiv \left(X_i^W, X_i^M\right)$, and $r_0\left(X_i\right) \equiv r_M\left(X_i^M\right) - r_W\left(X_i^W\right) + \Delta_0$. We assume that, while non-parametric with respect to X_i,[65] $r_0(\cdot)$ however satisfies the exclusion restriction that it does not depend on s_0^W, so that the gender bias in hiring is not itself a function of the share of women among all hired scientists. The error term ϵ_{0i} is distributed according to F, an unknown c.d.f. assumed to be invertible. Given these, the share of female scientists, 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 worldwide scale over the 20th century, we assume $X_i = X_0$ for all i's, with X_0 some constant value such that $r_M\left(X_0^M\right) = r_W\left(X_0^W\right)$.[66] 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 scientists.[67] 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 model, we allow the gender bias in publications Δ_1 to

[65] For example, $r_0(\cdot)$ trivially fits the linear specification $r_0\left(X_i\right) = X_i\beta_0 + \Delta_0$ assumed in equation (5), but can also take the more general form $r_0\left(X_i\right) = h\left(X_i, \beta_0\right) + \Delta_0$, with $h\left(X_i, \beta_0\right)$ any function of X_i and the parameter β_0 (of any dimension).

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

[67] 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 $r_0\left(X_0, s_0^W\right) = r_0\left(s_0^W\right)$, equation (C.4) would still take the rather inconvenient form $s_0^W = 1 - F\left(r_0\left(s_0^W\right)\right)$.

be a function of the share of female scientists, $\Delta_1\left(s_0^W\right)$. 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:

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

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.

Publications Conditional on Gender

The expectation of Y_{1i}^W conditional on $Y_{0i} > 0$ from equation (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 equation (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 not possible to separately identify the various components of $g_W(\cdot)$ and $G_M(\cdot)$.[68] To avoid unnecessarily strong functional form restrictions, we directly approximate these functions by polynomial expansions of the share of female scientists: $g_W\left(s_0^W\right) = \sum_{\kappa=0}^{2} \theta_\kappa^W \times \left(s_0^W\right)^\kappa$ and $G_M\left(s_0^W\right) = \sum_{\kappa=0}^{2} \theta_\kappa^M \times \left(s_0^W\right)^\kappa$, respectively. Following the same steps used to obtain regression (12), by assuming that $Z_i^W = Z_i^M = Z_i$, we estimate equations (C.6) and (C.7) on the basis of a the following regression:

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

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

where Pub_{it} measures the standardized number of papers published by scientist i in cohort $t(i)$, $s^W_{0\ell(i)}$ is the share of female scientists in i's cohort-country $\ell(i)$ (e.g., the USA in 2000), and each $\gamma_{\kappa p}$ corresponds to the portion of gender gap $\theta^W_{\kappa p} - \theta^M_{\kappa p}$, $\kappa = 0, 1, 2$, which we estimate separately for each period $p = 1, 2, 3$.

The estimation results of this version of the model are reported in the main text in Table 4, columns (4)-(6), and the corresponding predicted gender gap functions plotted in Figure 7, panel (b). The figure summarizes the estimation results by plotting, separately for each period p, the predicted gender gap function $\widehat{g}_p\left(s^W_{0\ell}\right) = \widehat{g}_{W,p}\left(s^W_{0\ell}\right) - \widehat{G}_{M,p}\left(s^W_{0\ell}\right) = \sum_{\kappa=0}^{2} \widehat{\gamma}_{\kappa,p} \times \left(s^W_{0\ell}\right)^{\kappa}$ against observed values of $s^W_{0\ell}$.

D Further Details and Results: Citation Gaps

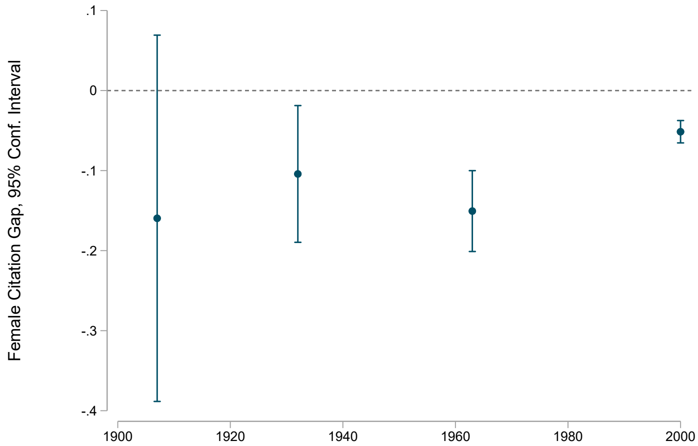
D.1. Further Details: Predicted Citations

As outlined in the main text, we aim to account flexibly for the topic of each paper which could influence citations. We estimate a ridge regression that uses the words (unigrams) and word pairs (bigrams) that appear in the title of the 749,291 scientific papers, that we match to at least one scientist in our data. The model learns how many citations, on average, papers in finely-grained research fields typically receive.[69]

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 plus 30 indicators for the length of titles.[70]

The model minimizes equation (13) 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.[71] 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 disciplines. The model can thus account for the changing importance of topics over time and across disciplines.

Figure D.1: Gender Gaps in Citations over Time



Notes: The Figure shows gender gaps in citations over time for the sample of prestigious universities (sample 3). The gender gaps are estimated with equation (14). The regression controls for cohort-discipline-country fixed effects and the predicted citation controls.

D.2. Further Results: Citation Gaps

Table D.1: Citations Gaps: Controlling for Predicted Citations (Robustness)

	Baseline (1)	Out of Sample (2)	Out of Sample (3)	Linear Control (4)	Non-Parametric Control (5)	Without Winsorization (6)	Citation Count (7)	Double ML (8)
Sample 2: All Universities, Sciences, 1900-1969								
Female-First/Last Author (1900/14)	-0.143*	-0.116	-0.093	-0.160*	-0.142*	-0.139***	-2.975**	-0.103*
	(0.082)	(0.100)	(0.090)	(0.086)	(0.079)	(0.032)	(1.377)	(0.054)
	[0.132]	[0.134]	[0.116]	[0.142]	[0.143]	[0.093]	[3.278]	
Female-First/Last Author (1925/38)	-0.130***	-0.134***	-0.131***	-0.159***	-0.132***	-0.075***	-3.310*	-0.122***
	(0.035)	(0.038)	(0.038)	(0.033)	(0.034)	(0.028)	(1.902)	(0.033)
	[0.051]	[0.044]	[0.048]	[0.05]	[0.049]	[0.035]	[2.711]	
Female-First/Last Author (1956/69)	-0.124***	-0.118***	-0.114***	-0.132***	-0.118***	-0.081***	-7.318***	-0.124***
	(0.017)	(0.015)	(0.013)	(0.018)	(0.017)	(0.013)	(1.033)	(0.013)
	[0.025]	[0.019]	[0.021]	[0.025]	[0.028]	[0.019]	[1.5]	
Observations	255,768	255,768	255,768	255,768	255,768	255,768	255,768	255,768
R^2	0.470	0.097	0.098	0.433	0.504	0.547	0.678	
Sample 3: Prestigious Universities, Sciences, 1900-2000								
Female-First/Last Author (1900/14)	-0.160	-0.143	-0.122	-0.161	-0.171	-0.102	-3.879*	-0.116
	(0.117)	(0.130)	(0.110)	(0.118)	(0.114)	(0.063)	(2.020)	(0.084)
	[0.123]	[0.143]	[0.145]	[0.123]	[0.125]	[0.082]	[3.779]	
Female-First/Last Author (1925/38)	-0.104**	-0.101**	-0.101**	-0.125***	-0.106**	-0.054	-1.807	-0.095**
	(0.044)	(0.044)	(0.044)	(0.040)	(0.043)	(0.043)	(2.951)	(0.045)
	[0.06]	[0.053]	[0.06]	[0.058]	[0.06]	[0.048]	[3.6]	
Female-First/Last Author (1956/69)	-0.151***	-0.144***	-0.142***	-0.152***	-0.149***	-0.100***	-8.949***	-0.144***
	(0.026)	(0.024)	(0.023)	(0.026)	(0.026)	(0.018)	(1.728)	(0.019)
	[0.033]	[0.03]	[0.034]	[0.033]	[0.033]	[0.025]	[2.134]	
Female-First/Last Author (2000)	-0.051***	-0.063***	-0.064***	-0.056***	-0.053***	-0.043***	-2.824***	-0.068***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.426)	(0.005)
	[0.009]	[0.009]	[0.011]	[0.009]	[0.009]	[0.007]	[0.449]	
Observations	611,513	611,513	611,513	611,513	611,513	611,513	611,513	611,513
R^2	0.375	0.133	0.137	0.356	0.395	0.394	0.641	
Cohort×Discipline×Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predicted Citations Control	Yes			Yes		Yes	Yes	
Predicted Citations Control (Out-of-Sample)		Yes						
Predicted Citations Control (All, Out-of-Sample)			Yes					
Predicted Citations (1000 bins) FE					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. T The main explanatory variable is an indicator that equals 1 if the paper's first or last author is a woman, interacted with the relevant cohort(s). The regressions control for 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 indicators for the permilles of the predicted citation distribution interacted with discipline indicators. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model. Standard errors are clustered at the discipline-country level with 781 clusters in sample 2 and 1,816 in sample 3. We additionally report boostrapped standard errors in square brackets. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

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

[70] The last indicator equals one for all titles with 30 words or more.

[71] We consider values of λ in a range of 1 to 10.

Table D.2: Citations Gaps: Definition of Female Paper (Robustness)

	Baseline (1)	Female Indicator (2)	Female First-Author (3)	Female Share (4)
Sample 2: All Universities, Sciences, 1900-1969				
Female Paper (1900/14)	-0.143*	-0.145*	-0.138*	-0.136
	(0.082)	(0.080)	(0.080)	(0.090)
	[0.132]	[0.134]	[0.116]	[0.142]
Female Paper (1925/38)	-0.130***	-0.104**	-0.136***	-0.107***
	(0.035)	(0.044)	(0.046)	(0.041)
	[0.051]	[0.044]	[0.048]	[0.05]
Female Paper (1956/69)	-0.124***	-0.115***	-0.133***	-0.130***
	(0.017)	(0.017)	(0.023)	(0.018)
	[0.025]	[0.019]	[0.021]	[0.025]
Observations	255,768	255,768	255,768	255,768
R^2	0.470	0.469	0.469	0.470
Sample 3: Prestigious Universities, Sciences, 1900-2000				
Female Paper (1900/14)	-0.160	-0.161	-0.153*	-0.161
	(0.117)	(0.115)	(0.085)	(0.116)
	[0.123]	[0.143]	[0.145]	[0.123]
Female Paper (1925/38)	-0.104**	-0.066	-0.115**	-0.068
	(0.044)	(0.057)	(0.055)	(0.058)
	[0.06]	[0.053]	[0.06]	[0.058]
Female Paper (1956/69)	-0.151***	-0.150***	-0.165***	-0.170***
	(0.026)	(0.024)	(0.028)	(0.028)
	[0.033]	[0.03]	[0.034]	[0.033]
Female Paper (2000)	-0.051***	-0.008	-0.058***	-0.028***
	(0.007)	(0.007)	(0.010)	(0.009)
	[0.009]	[0.009]	[0.011]	[0.009]
Observations	611,513	611,513	611,513	611,513
R^2	0.375	0.375	0.375	0.375
Predicted Citations Control	Yes	Yes	Yes	Yes
Cohort×Discipline×Country FE	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 variables are different definitions of female-authored papers, interacted with the relevant cohort(s). Column 1 uses an indicator that equals 1 if the paper's first or last author is a woman. In column 2, the indicator variable is equal to 1 if the paper has at least 1 female author. In column 3, the indicator variable is equal to 1 if the first author of the paper is a woman. Column 4 uses the share of female authors on a paper. 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. Standard errors are clustered at the discipline-country level with 781 clusters in sample 2 and 1,816 in sample 3. We additionally report bootstrap standard errors in square brackets. Significance levels: *** p<0.01, ** p<0.05, and * p<0.1.

E Further Results: Promotion Gaps

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

Promotion to Full Professor		*Sample 2*			*Sample 3*	
	Men	Women	Total	Men	Women	Total
No Promotion	72.01%	88.04%	73.13%	70.18%	85.38%	70.95%
Promotion in Worse Department	4.73%	1.03%	4.47%	5.40%	1.73%	5.21%
Promotion in Same Department	19.87%	10.07%	19.18%	20.32%	12.11%	19.90%
Promotion in Better Department	3.39%	0.86%	3.22%	4.10%	0.79%	3.93%
Observations	15,411	1,162	16,573	11,944	636	12,580

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