Gender Gaps in Academia: Global Evidence Over the Twentieth Century

Alessandro Iaria (University of Bristol)

Carlo Schwarz (Bocconi University)

Fabian Waldinger (University of Munich)*

December 9, 2021

Abstract

We analyze gender gaps in academia along four dimensions (hiring, publications, citations, promotions) over an unprecedented time-span and geographic coverage. For this analysis, we hand-collect the largest database of university academics ever assembled. First, we document that the world-wide share of women rose from 1% in 1900 to 11% in 1969. Anglo-Saxon countries, were at the vanguard of hiring women, while Germanic countries lagged behind. Second, we estimate negative gender gaps in publications of about 0.2 sd., which do not narrow over time. The publication gap is positive in countries and periods with very low female shares but turns negative yet narrowing with increasing female shares. Third, we estimate negative gender gaps in citations which hold controlling for the topics of papers with a novel machine-learning approach. Fourth, we show negative gender gaps in promotions which hold controlling for publication records.

*We thank Matthias Bing, Zishan Bhatti, Angela Buensuceso, 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, Robert Johannes, Simon Kallmaier, Christoph Kuehn, Axel-Konstantin Marko, Lukas Mergele, Luisa Neumeier, Tobias Nowacki, Adrian Nutzinger, Franz Oertel, Julia Ostendorf, Nicolas Porschke, Joseph Prestwich, Akos Reitz, Tabea Rudolf, Riyan Sawalha, Constanze Schwarz, 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 also thank David Card, Amy Finkelstein, Claudia Goldin, Ben Jones, Claudia Olivetti, Nina Roussille and seminar participants at Berkeley, Bocconi, Boston University, Boulder, Bristol, Frankfurt, Munich, NYU, Turin, UCLA, Warwick, and York for useful comments and suggestions. We also thank Martin Watzinger for sharing the Microsoft Academic Graph data. Waldinger gratefully acknowledges startup funding from ERC Starting Grant N°335573.

1 Introduction

Until the beginning of the 20th century, high-skilled professions were almost exclusively occupied by men. Even today, women remain under-represented in high-skilled professions, especially in senior positions. For example, in 2017 only 19.9% of corporate board seats and 5.8% of CEO positions in U.S. Fortune 500 firms were held by women (Bertrand et al., 2019), and only 9.5% of inventors in OECD countries were women (OECD, 2021).

This paper is the first attempt to measure the evolution of gender gaps in a high-skilled profession over a large part of the twentieth century at a global scale. For this purpose, we hand-collect the largest database of university academics ever assembled. The data contain faculty rosters in 7,484 universities for all academic disciplines in more than 130 countries for six cross-sections (cohorts) covering the years 1900, 1914, 1925, 1938, 1956, and 1969. For comparison, the U.S. News ranking of world-wide universities is based on a pool of 1,748 universities (see USNEWS/Methodology, accessed August 6, 2021). The Shanghai Ranking of World Universities includes 2,417 universities (see https://www.shanghairanking.com/institution, accessed August 6, 2021).

We make a large number of manual enhancements to the faculty rosters. First, we code the gender of academics. Second, we follow academic careers across the six cohorts using a cascading merge procedure. For example, we trace Margarete Bieber's career from the University of Gießen, Germany (1925 cohort), to the Columbia University, USA (1938 and 1956 cohorts).[1] Third, we manually recode more than 100,000 specializations into 36 disciplines. For example, the specializations "advanced reactor theory and quantum theory" or "physique des particules élémentaires" are assigned to "physics." Fourth, we consistently code academic ranks across countries, e.g. professor, associate professor, assistant professor. Finally, we enrich the dataset with publication and citation data from the *Clarivate Web of Science* and *Microsoft Academic Graph*.[2]

Crucially, the data cover almost all academics and not only the tiny sliver of the most prominent ones. Because data on prominent academics are more readily available, the history of science has been told predominantly as a tale of male star scientists in a small number of countries and elite universities. Moreover, the data are based on faculty rosters and not on publication databases. This enables us to study the whole population at risk of publishing or of receiving a promotion to the highest academic rank. These features help us overcome important selection biases and depict a more complete picture of the role of women

[1] Margarete Bieber was an archaeologist and art historian and became the second woman in Germany to become full professor at a university. As she was Jewish she was dismissed from her post by the Nazi government and emigrated to the United States.

[2] As always with data on a global scale, there can be concerns about the comparability across countries and over time. We discuss these in Section 2 and in the Appendix.

in academia. Based on these data, we establish new facts about gender gaps in academia along four dimensions: hiring, publications, citations, and promotions.

First, we document gender gaps in hiring. We show that the share of women in all universities across the globe combined was only 1 percent in 1900.[3] In the following decades, the share of women increased slowly: 2 percent by 1914, 3 percent by 1925, 7 percent by 1938, and 11 percent by 1956 and 11 percent by 1969. We further investigate how gender gaps in hiring vary across academic ranks and document even larger gaps among full professors. By 1900, all universities across the globe combined only hired 111 women as full professors, again a share of around 1 percent. In the following decades, the share of women among full professors increased, but remained always below the share among all academics. The slower increase in the share of women among full professors, compared to all academics, could either reflect compositional changes over time or worse career prospects for female academics. We investigate this question at the end of the paper.

The global coverage of the data also enables us to explore heterogeneity in the share of female academics across countries. We provide new evidence that the United States played *the* leading role in hiring female academics: universities in the United States hired more female academics than the rest of the world combined. Overall, more than 70 percent of female academics were employed in the United States. The share of female academics in the United States was almost twice as high as in any other country. More generally, the share of female academics was particularly high in Anglo-Saxon countries. On the other side of the spectrum, the share of female academics was particularly low in countries which organized their university systems in the German tradition (e.g., Germany, Austria, Finland, or Sweden).

Next, we document substantial heterogeneity in female shares across disciplines. Averaged over the period 1900 to 1969, no discipline had a female share greater than 35 percent. Most disciplines had female shares below 10 percent. Disciplines with particularly high female shares were communication studies, sports sciences, social sciences, and natural sciences.[4] Most of these disciplines were not research-oriented but rather focused on teaching to undergraduates. We document female shares below 2 percent, averaged over the entire time period, in veterinary medicine, architecture, theology, engineering, and law.

[3] In all universities covered by our data only 228 women had been hired in 1900. For many disciplines, countries, and universities the data cover the first women to enter academia and the first woman promoted to full professor, e.g. Katherine Ellis Coman, the first female full professor in economics in the United States (Vaughn 2004).

[4] Academics are only classified as natural scientists if they report "natural science" as their specialization. Academics who report a specialized discipline, e.g. physics or chemistry, are not classified as natural scientists but as physicists or chemists. Similarly, academics are only classified as social scientists if they do not report a specialized discipline within the social sciences.

Second, we investigate gender gaps in academic output, as measured by publications. One of the unique advantages of studying academics is the availability of output measures that are comparable across time and space. Publications are key performance metrics that have been widely used to evaluate academics. It is important to note that, of course, publications do not measure the true ability of an academic. They reflect realized output that can be affected by preferences, discrimination, and other biases. Because our data collection relies on complete faculty rosters, we observe all academics independently of whether they publish. This helps to overcome important selection concerns when comparing publication output across gender. We measure publications over a $\pm$ 5 year interval around each cohort (e.g., 1909 to 1919 for scientists observed in 1914). We show that female academics publish 2.5 fewer papers than their male peers, which corresponds to around 0.22 standard deviations. The gender gap in publications is similar if we compare men and women in the same department and cohort, e.g. physics in Harvard in 1969. Strikingly, we do not find evidence of a narrowing publication gap over the 70 years covered by our data.

We also investigate the relationship between the share of female academics and the publication gap in a country. We find a (Nike) "swoosh"-shaped relationship. This "gender-swoosh" suggests that publication gaps were mostly positive in countries and periods with very low shares of female scientists. This can be referred to as the "Marie-Curie" period: only exceptional women were hired and, on average, published more than men despite potential discrimination and other biases in the publication market.[5] With increasing shares of women in the profession, gender gaps in publications turned negative. However, when the share of women increased beyond very low levels, the negative gender gaps in publications narrowed. We outline a model along the lines of Roy (1951) to interpret this relationship. The proposed model allows for (i) selection on unobservables in the hiring market, (ii) gender bias in hiring, and (iii) gender bias in the publication market. These features lead a scientist's publication outcome to be a function of the share of women in a country-period because of (a) *indirect* effects of selection and gender bias in the hiring market and (b) the *direct* effects of gender bias in the publication market. Based on the proposed model, we estimate a semiparametric regression which rationalizes the "gender swoosh." The results suggest that gender biases in hiring may have indirect repercussions on the observed productivity of female scientists.

Third, we explore gender gaps in citations and estimate whether papers published by women receive fewer citations. Importantly, we study if a potential citation gap stems from differences in topics that women work on. For this purpose, we develop a novel machine learning approach that allows us to study whether citation gaps can be explained by women

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

working on topics with a lower citation potential. We train a regularized regression model that uses the words in paper titles to predict the expected number of citations for each paper. To prevent the model from internalizing biases against papers published by women, the training sample solely consists of papers published by men. The method therefore allows to predict the citations of each paper, as if it had been written by men. Our results indicate that papers written by female authors receive fewer citations than papers by male authors. The gender gap of around 2 percentiles of the citation distribution holds even after flexibly controlling for the predicted citations of the paper. This suggests that citation gaps are not driven by women publishing on topics that generally receive fewer citations.

Fourth, we investigate gender gaps in promotions to full professor. We find that women are around 20 percentage points less likely to be promoted than men. This result holds even if we compare men and women who enter the data in the same department and cohort (e.g., physics in Berkeley in 1900) and if we control for the publication and citation record of the scientist. Strikingly, the unexplained gender gap in promotions is larger than the effect of a four standard deviations worse publication record.

Taken together, we show that there were significant gender gaps in hiring, publications, citations, and promotions. Such barriers that excluded women from participating in science may result in "lost Marie Curies."[6] Furthermore, gaps in recognition of their work not only impedes women's scientific careers but also deprives the scientific community of ideas and scientific breakthroughs. In a world where ideas play an ever-increasing role, this will slow down scientific progress and ultimately economic growth (e.g. Romer, 1986; Romer, 1990; Jones, 1995a).[7]

The paper contributes to a growing literature on gender gaps in science and innovation. In economics, female-authored papers receive more citations than similar male-authored papers, suggesting that women need to pass higher hurdles to publish (Card et al., 2020a); women were less likely to be nominated as Fellows for the Econometric Society until the late 1970s but more likely to be nominated since the mid-2000s (Card et al., 2020b); women receive less credit for group work (Sarsons, 2017; Sarsons et al., 2021); references to female-authored economics papers are more likely to be omitted (Koffi, 2021); and female-authored papers have higher readability scores (Hengel, 2020). Covering subjects beyond economics, the literature has also documented that US-based female scientists 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 volunteer and to be asked to volunteer for tasks with low

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

[7]Hsieh et al. (2019) use a structural Roy model to argue that declining discrimination against women and Blacks raised U.S. aggregate productivity.

promotability (Babcock et al., 2017); that a higher share of women in evaluation committees lowers promotion prospects of female academics in Italy and Spain (Bagues, Sylos-Labini, and Zinovyeva, 2017); and that physicians become more pessimistic about female surgeons' ability than male surgeons' ability after a patient's death (Sarsons, 2019). Compared to this earlier work, our contribution lies in providing a comprehensive analysis of gender gaps in academia, along four dimensions (hiring, publications, citations, and promotions), covering academics in *all* disciplines, in a large number of countries, and over a large part of the twentieth century.

Our work also contributes to the growing literature that analyzes gender gaps in certain high-skilled professions focusing on individual countries and time periods, e.g., MBA graduates (Bertrand, Goldin, and Katz, 2010), executives (e.g., Bertrand and Hallock, 2001; Gayle, Golan, and Miller, 2012; Albanesi, Olivetti, and Prados 2015), lawyers (Azmat and Ferrer, 2017), pharmacists (Goldin and Katz, 2016), and engineers (Roussille, 2020); all in the United States, or more broadly college graduates in the United States (Black et al., 2008) or Sweden (Albrecht, Björklund, and Vroman, 2003). Our new database enables us to trace the evolution of gender gaps for one high-skilled profession at the global level and over seven decades. In contrast, most existing papers have analyzed one country and relatively limited time periods because of a lack of comparable data.[8] A nuanced description 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 to overcome barriers that deprive academia, and society, of some of the best minds and ideas.

2 A New Database of University Academics

2.1 Hand-collecting Historical Faculty Rosters

At the heart of this paper is the largest database of university academics ever assembled. We hand-collect this database from the historical publication *Minerva Jahrbuch der Gelehrten Welt*. In a time before the Internet, *Minerva* was the most important world-wide direc-

[8]Another related literature has documented declining gender gaps in graduating from college in the United States (e.g. Goldin, Katz, and Kuziemko, 2006). The paper also contributes to the empirical literature on gender gaps. A large literature has studied gender gaps in hiring probabilities and wages (see Altonji and Blank 1999, Bertrand, 2011, Blau and Kahn, 2017, and Bertrand and Duflo, 2017 for extensive surveys). A notable exception regarding the time period is Claudia Goldin's seminal research on gender gaps in wages and employment in the United States from the late 19th century until today (e.g., Goldin, 1989, 1990).

tory of academics. The publishers of *Minerva* contacted ministries of education, university administrators, and academics to ensure an almost comprehensive coverage.[9]

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 the left side of Figure 2 for a sample page).[10] For the remainder of the paper, we refer to these years as cohorts. *Minerva* lists academics from all disciplines, thousands of universities in more than 100 countries. The data include traditional universities such as *Harvard* or the *University of Tokyo*, technical universities such as *MIT* or *École Polytechnique*, mining universities such as *Freiberg Mining Academy*, and theological universities such as *Pontificia Università Gregoriana in Collegio* in Rome. Importantly, virtually all Ph.D. granting institutions are included the data. For example, the data contain academics in 1,575 universities for the United States, 305 universities in the United Kingdom, 318 in Germany, and 348 in France.[11]

Minerva usually lists the name of the university as well as the city and country, followed by faculty rosters.[12] For most universities, the data list assistant, associate, and full professors, but also honorary professors, and in some cases research positions, and teaching positions. The faculty rosters usually report the name of each academic as well as a finely grained specialization. Overall, the data contain around half a million person-cohort observations (Table 1) in 7,484 universities in more than 100 countries.

The number of academics increased from around 25,000 in 1900 to close to 200,000 in 1969, reflecting the spectacular growth of the university sector during this period (Figure 1).

[9]For example, an article in *Nature* compared the French publication *Index Generalis: Annuaire Général des Universités* to *Minerva* and noted that "[i]n scope, as indicated by the sub-title, this annual is akin to the well-known 'Minerva Jahrbuch der gelehrten Welt'. It is, however, very much less exhaustive" (Nat 1930). To the best of our knowledge, there are no comparable data that cover academics on a world-wide scale over many decades. To provide evidence of its coverage, we benchmark the *Minerva* data to smaller datasets that cover some universities and time periods. The benchmarking exercises suggest that *Minerva* indeed covered a large fraction of the world's academics (see Appendix 8.4 for details).

[10]As the number of universities and academics greatly increased over time, *Minerva* published the data for the last two cohorts in two installments. We refer to these cohorts using the later year, i.e. 1956 for the 1952/56 publication.

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

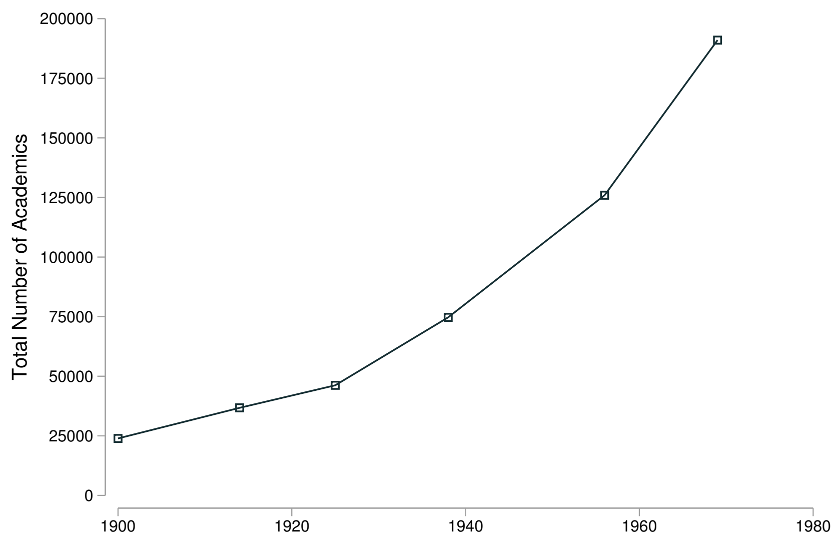
[12]For some less known universities, especially in India, the source only reports the number of professors without listing their names. Furthermore, for some universities the source lists the names of professors but only reports the number of teaching positions (e.g. "10 lecturers") without listing names. Across all cohorts, the source list 498,528 faculty members with names (this sample forms the basis for our analysis, Table 1) and 108,398 additional faculty members (e.g. the 10 lecturers) without names.

Table 1: Summary Statistics

	Full Sample				Scientist Sample with Gender		
	All	All Gender Coded	Female	Male	All	Female	Male
Number of universities	7,483	5,497	2,421	5,123	2,315	1,081	2,228
Number of departments	36,973	30,470	8,834	28,895	7,138	2,133	6,753
Number of academic - cohort observations	498,528	409,764	35,637	374,127	132,677	7,661	125,016
Female %		8.70	100.00	0.00	5.77	100.00	0.00
Publications					4.27	3.14	4.34

Notes: The Table shows summary statistics at the academic-cohort level. The scientist sample contains all academics in mathematics, physics, chemistry, biochemistry, biology, and medicine. The data were collected by the authors from various volumes of Minerva, see section 2 for details.

Figure 1: Number of Academics Over Time



Notes: The Figure shows the total number of academics in the six Minerva cohorts. The data were collected by the authors from various volumes of Minerva, see section 2 for details.

Digitization and Manual Enhancements of Academics Data

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

In addition, we make a large number of manual enhancements that involve extensive hand-checking (see Appendix 8.1 for details on each of the steps reported below). First, we manually re-code over 4,000 different university ranks (e.g., “professor,” “chargé de cours,” or “incaricato”) into ten comparable categories (e.g., assistant professor, full professor, emerita/us, or teaching position). Second, we manually re-code over 100,000 different specializations (e.g., "Advanced Reactor Theory and Quantum Theory" or "Physique des particules élémentaires") into 35 disciplines (e.g., physics, economics, law, theology, or history). Third, if academics hold multiple positions within the same city or university (e.g., a double appointment in two departments), we combine the information into a single observation. Fourth, we link academics across cohorts using a cascading linking procedure. In total, we are able to link 168,090 person-cohort observations across cohorts.[13] Fifth, for academics who are only listed their surname and initials (instead of the complete first name) we conduct a manual web search to find their complete first name.[14] Sixth, we obtain consistent university identifiers by linking universities across cohorts and tracking mergers and splits of universities.

Identifying the Gender of Academics

We develop a new five-step procedure to identify the gender of academics at a global scale. First, whenever available, we use information on gender from the faculty rosters in *Minerva* (e.g., names preceded by Mlle., Lady, Lord, Cardinal, and so on). In all further steps, we rely on first names to identify the gender of academics.

In the second step, we process more than 100,000 'first name'-country combinations with *gender-api.com*, a commercial solution that allows to code gender on the basis of first names and countries.[15] *Gender-api.com* assigns a gender probability to 'first name'-country combinations.

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

In the fourth step, we process the remaining cases that *gender-api.com* classified as less than 100% male by googling the 'first name'-country combination using a *Google* picture search. A research assistant then classifies the 'first name'-country combination as male or female depending on whether the picture search returns more male or female individuals. E.g., *gender-api* and the research assistants could not identify the gender of "Hadmar" in Austria. We therefore search for "Hadmar Austria" in *Google* and analyse the pictures that

[13]This is substantially lower than the total number of academics, because many of them first entered the data in the last cohort (1969) and, hence, cannot be linked over time.

[14]All results remain unchanged without this step.

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

Google returns. In this example, the pictures that depict individuals show only men (see Appendix Figure A.1), we therefore code the 'first name'-country combination as male.

In the fifth step, we hand-check individual academics who appear to be mis-classified with an extensive *Google* search.[16] Such mis-classifications mostly occur because the predominant gender of some first names changed over time. E.g., the French name "Camille" can be both male and female. In the early cohorts, most academics with the first name "Camille" are male, while in later cohorts some are female. While these manual steps significantly increase the data quality, none of the results change without steps 3 to 5. For our main results, we focus on the sample of 409,767 academics for whom we can assign gender (Table 1).

Examples of Academics in the Database

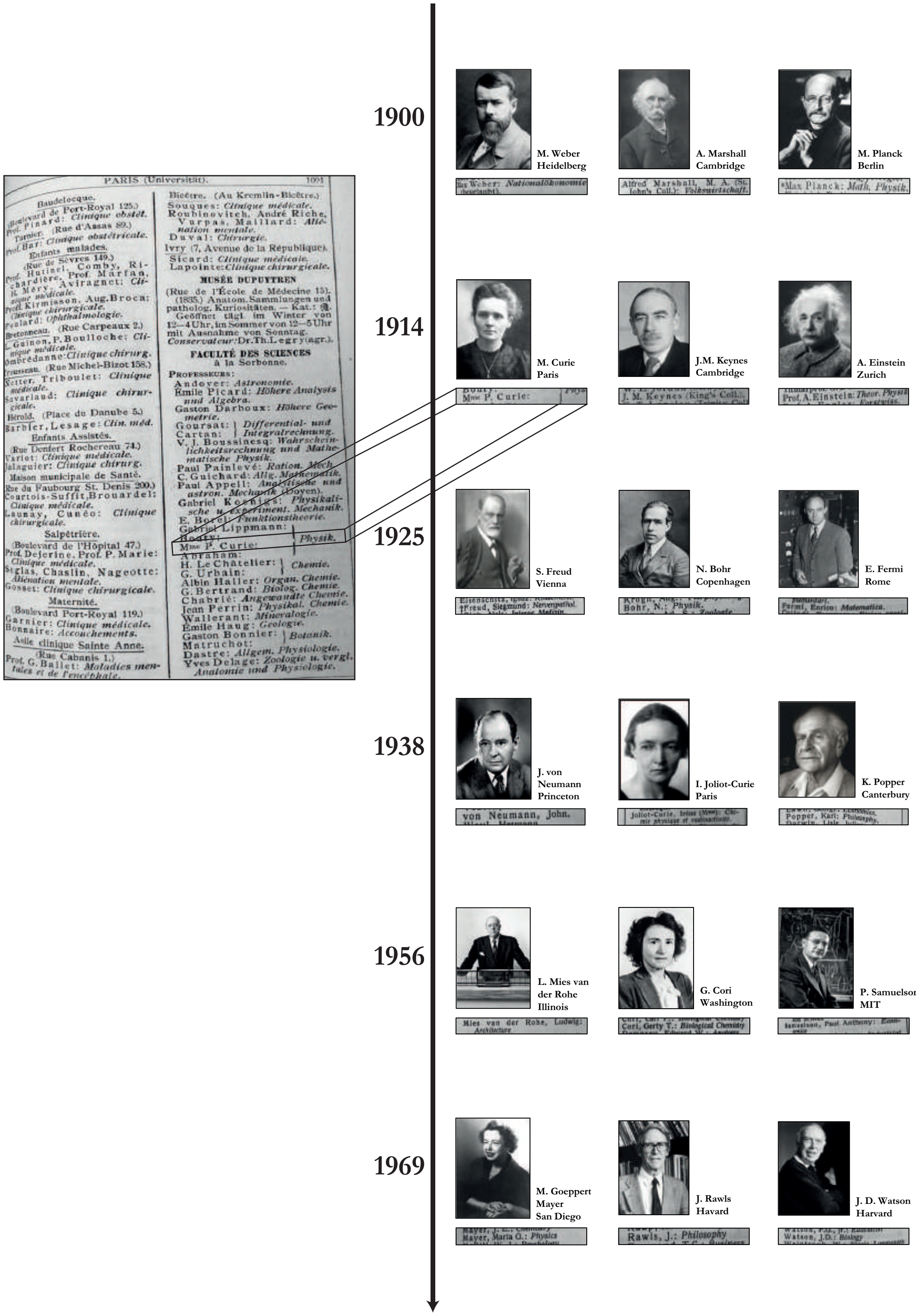
Figure 2 shows three exemplary academics for each of the cohorts. Choosing examples among the half a million academic-cohort observations leads to somewhat arbitrary decisions. The selection showcases some of the country, discipline, cohort, and gender dimensions of the data. However, it will not do justice to the tens of thousands of academics that have contributed to the progress of knowledge. For 1900, the data include the economist Alfred Marshall (University of Cambridge), the physicist and Nobel Laureate Max Planck (University of Berlin), and the sociologist Max Weber (University of Heidelberg).

Examples for 1914 are the economist John Maynard Keynes (University of Cambridge), the physicist and Nobel Laureate Albert Einstein (ETH Zürich), and arguably the most famous female academic in our data, Marie Curie (Université de Paris). Together with her husband Pierre Curie she conducted pioneering research on radioactivity and was the first woman to win the physics Nobel prize in 1903. Despite this achievement, she was not awarded a professorship at the Sorbonne. Only after her husband had tragically died, she finally became the first female full professor at the Sorbonne, five years after winning her first Nobel Prize and two years before she won her second, this time for her contributions to chemistry (McGrayne, 1998).[17]

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

[17]During her entire career, Marie Curie faced obstacles because of her gender. In 1911, she was nominated to the French Academy of Sciences. Her nomination met strong resistance: "Women cannot be part of the Institute of France" argued the physicist Émile Amagat. Despite the efforts of some of France's greatest scientists (e.g., Poincaré, Picard, or Lippmann), she lost the membership election by one vote to her male competitor (see Curie, 1938, pp. 277 for a detailed account). "She never again sought membership in the Academy ... As for the Academy, it refused to admit women until 1979." (McGrayne, 1998, p. 30).

Figure 2: Examples of Academics in Database



Notes: The Figure provides three examples of notable academics for each of the six Minerva cohorts. Strikingly, Marie Curie is listed as Mme P[ierre] Curie, at a time when she had won two Nobel prizes and her husband had tragically died in a road accident). Most of Marie Curie's papers were published under the name Mme P. Curie.

For 1925 the data contain the physics Nobel Laureate Niels Bohr (University of Copenhagen), the founder of psychoanalysis Siegmund Freud (University of Vienna) and the physicist Lise Meitner (University of Berlin). Lise Meitner was born in Vienna and became the second woman to earn a PhD in physics at the University of Vienna. During her Post-Doc years at the University of Berlin she did not receive a salary and had to run her experiments in a converted carpenter's shop in the basement because — as a woman — she was not allowed to enter the main building of the laboratory. During the 1920s she made important contributions with her collaborator Otto Hahn. The Nobel Laureate Wolfgang Paul commented "Hahn and Meitner were great friends, but when they talked, she was superior." In 1945, the Swedish Academy awarded the Nobel Prize to Otto Hahn but overlooked Lise Meitner's contribution to the Nobel Prize winning work. In the words of Ernst Fischer the decision to omit Meitner is a "stupidity of the Swedish Academy" (Kricheldorf 2014, p. 219).

Examples for 1938 are the mathematician John von Neumann (IAS Princeton) and the philosopher of science Karl Popper (University College Canterbury, NZ). Both of them fled Nazi persecution and had emigrated by 1938 (see e.g. Becker et al., 2021). Another example for 1938 is Irène Joliot-Curie, Pierre and Marie Curie's daughter and only the second women 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). Gerty Cori was the first woman to win the Nobel prize in physiology/medicine in 1947 (and the third woman to win a science Nobel prize). Despite her talent, Cornell, Toronto, and Rochester refused to hire her while offering professorships to her husband, and fellow Nobel Laureate, Carl Cori. In 1931, Washington University made both of them an offer, but Gerty was hired as a research associate while Carl was hired as full professor. Her salary was 20 percent of Carl's (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 molecule James Dewey 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 the University of California, San Diego (Wigner, 1972). In 1963, she became the second woman to win the Nobel Prize in physics, 60 years after Marie Curie.

2.2 Publication and Citation Data

To study gender gaps in publications and citations, we augment the *Minerva* data with publication and citation data from *Clarivate Analytics Web of Science*. For any result based on publications and citations, we focus on five academic disciplines: medicine, biology, chemistry, physics, and mathematics, which cover around a third of all academics in our data. We refer to this sample as the scientist sample. It contains 132,677 academic-cohort observations. The sciences have particularly good coverage in the *Web of* Science. By 1900, these disciplines had already established a culture of publishing in scientific journals and the publishing process was similar to today's. Furthermore, the publishing process was highly international and many scientists published in international academic journals (see Iaria, Schwarz, and Waldinger, 2018). For the years of our study, the *Web of Science* contains papers in 3,864 journals. Naturally, the coverage of the *Web of Science* is not uniform across countries and over time. This does not affect our estimates of gender gaps because we control for time and country (or even finer) fixed effects in all regressions reported below.

We match publications to the academics using a cascading matching algorithm (see Appendix 8.3 for details). The matching is based on the academic's surname, first name or initials (depending on whether first names are available), country, city, and discipline.[18] To harmonize addresses across *Minerva* and the *Web of Science* (and within the *Web of Science*), we process detailed addresses, e.g., "Cavendish Laboratory, Cambridge University, UK", with *Google Maps API*. This allows us to extract cities and countries for each of the hundreds of thousands of relatively unstructured addresses. E.g., we extract the city 'Cambridge' and the country 'United Kingdom' for the 'Cavendish Laboratory' address.

The matching of academics to publications is always based on the main discipline of the academic (e.g., physics) to reduce the number of false positives. As the *Web of Science* only assigns disciplines (e.g., physics, chemistry, or general science) at the journal-level, we develop a machine-learning classifier to assign disciplines to individual papers. Assigning the correct discipline is especially important for papers published in multi-disciplinary journals, such as *Nature* or *Science*. These papers could otherwise not be merged on the basis of a discipline. The classifier is based on a L2-regularized multinomial logit model. The model predicts a

[18]For the time period we study, the *Web of Science* only reports initials of authors for most papers. Sometimes, the *Web of Science* does not report addresses, even though the original paper actual lists an address. In some of these cases an alternative database, *Microsoft Academic Graph* (*MAG*), contains the relevant address information. We therefore enrich the addresses with information from *MAG* (see Appendix 8.2.2 for details).

discipline for each paper, based on the unigrams, bigrams, and trigrams from the titles of the 59% papers which were published in journals that are assigned to only one discipline (e.g., the *Physical Review*). The classifier achieves an out-of-sample F1-score of 0.81.[19]

To each scientist-cohort observation, we match publications in a ± five-year-window around the year of the corresponding *Minerva* cohort. E.g., for scientists listed in *Minerva* 1914, we match papers published between 1909 and 1919.[20] In the rare cases that two or more scientists have identical names and work in the same discipline, we assign the paper proportionally to each scientist (see Appendix 8.3).[21]

As the resulting dataset is based on complete faculty rosters, we can observe academics who publish but also those who do not publish in journals that are covered by the *Web of Science*. In contrast, similar databased on publications alone would result in a selected sample that would bias estimation results. Our data construction helps to overcome such selection bias in the measurement of scientific output.

3 Gender Gaps in Hiring

Hiring of Women Over Time

In the first part of the analysis, we investigate the evolution of gender gaps in hiring. Our new data show that in 1900 only 228 women had been hired across all universities in our data, a share of about 1 percent (Figure 3). In the following decades, the share of women in academia increased, in particular between 1925 and 1938, i.e., before WWII.[22] By 1969, a

[19]The F1-score is defined as the harmonic mean of precision (the number of true positives divided by the sum of true positives and false positives) and recall (the number of true positives divided by the sum of true positives and false negatives).

[20]We use a ± five-year-window because scientists do not necessarily publish in every year and thus we prefer not to merge papers published in only in the year of the publication of the *Minerva* data. The results are very similar if we match publications and citations in a ± three-year-window. As we merge papers on the basis of the surname and first name/initials a possible concern is that women changed their surname after marriage. This would prevent us from merging all relevant papers to female scientists within the ± five-year-window. Reassuringly, the estimated gender gaps remain unchanged if we merge publications in shorter time windows, making name changes less likely. This supports the view that name changes do not substantially bias our results. The absence of bias may be explained by two factors. First, *Minerva* predominately lists academics who are assistant, associate, or full professors and, hence, individuals who were already married if they ever got married. Second, marriage rates for female academics in this period are relatively low, e.g. 18 percent in 1921 and 26 percent in 1938 for scientists in the United States (Rossiter, 1982, p. 140).

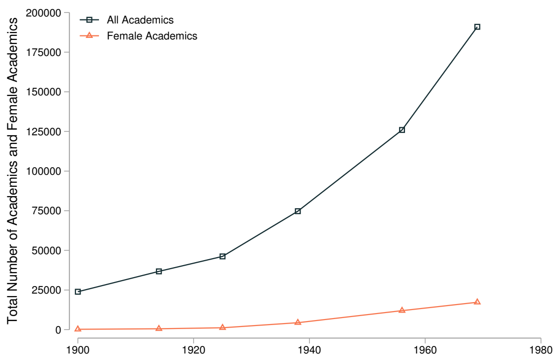
[21]Results are robust in a sample of scientists with unique lastname - first initial - discipline combinations in each cohort (i.e. for academics where *Minerva* only lists one academic with the same lastname, first initial, and discipline in any university of the world, see appendix Table C.1).

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

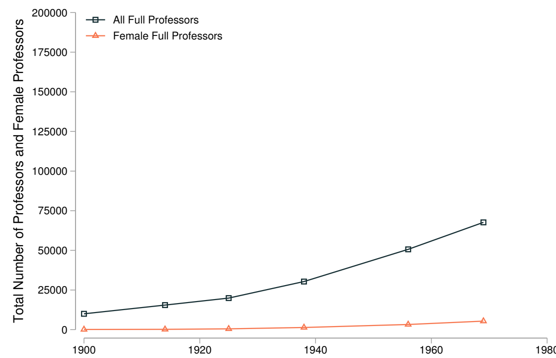
total of 17,276 women worked at the universities in our data. This corresponds to about 11 percent of all academics — still nowhere close to equal representation.

Figure 3: All Academics and Female Academics over Time

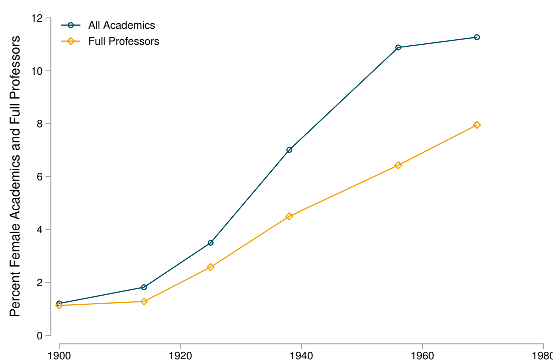
(a) All Ranks



(b) Full Professors and Female Full Professors



(c) Percent of Female Academics (All Ranks and Full Prof.)



Notes: The Figure shows the total number of academics and the total number of female academics in the six Minerva cohorts. Panel (a) shows academics of all ranks. Panel (b) shows full professors, only. Panel (c) shows the percentage of women among all academics and full professors, respectively. The data were collected by the authors from various volumes of Minerva, see section 2 for details.

We also explore changes in the number of women who held full professor positions. All over the world, full professor is the highest academic rank and professors have unique privileges and particularly high job security and salaries. Furthermore, full professor is the most comparable academic rank across the different university systems. In 1900, only 111 women worked as full professors across all universities in our data, representing about 1 percent of the profession. In the following decades, the share of women among full professors increased, and by 1969 reached about 8 percent.

The slower increase in the share of women among full professors, compared to all academics, may indicate that women were less likely to be promoted but may also reflect compositional changes, e.g., if a higher share of women was hired in later cohorts and the number of academics was increasing over time, it could take time for women to rise through the ranks. We systematically explore the role of gender for promotions to full professor in section 6.

Hiring Gaps Across Countries

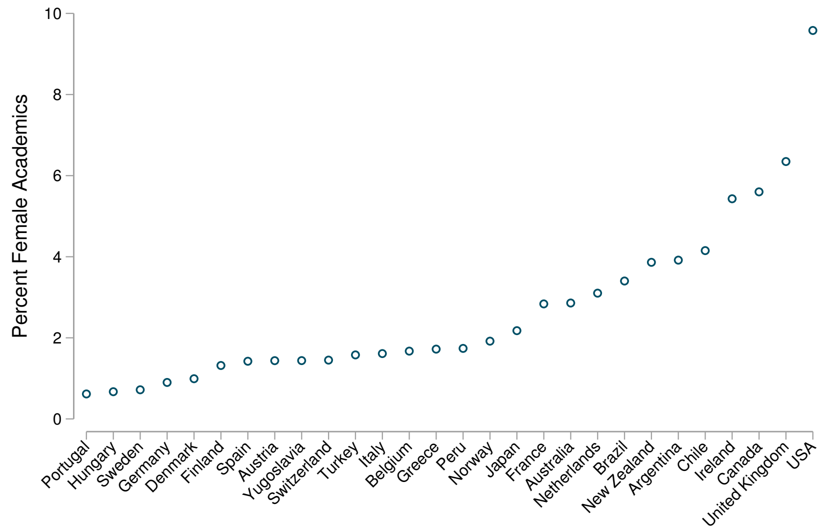
The overall increase in the number and the share of female academics masks significant heterogeneity across countries. The United States, in particular, stands out: already in 1900, more female academics worked in the United States than in all other countries of the world combined. The dominant role of the United States further increased until 1969. In that year, 70 percent of all female academics worked in the United States. Over the entire period, the average female share in the United States was around 10 percent (Figure 4, panel a).[23] Compared to the countries with the lowest share of female academics (Portugal, Hungary, Sweden), the United States had a 10-times higher share. Even compared to Canada, the country with the second highest share of female academics, the United States' share was nearly twice as high. These findings are qualitatively similar in a sample of universities that we observe in all six cohorts and in a sample that excludes women's colleges (see Appendix Figures B.2 and B.3). This suggests that neither compositional changes due to university entry nor the presence of women's colleges in the United States explain these findings.

More generally, Anglo-Saxon countries had higher shares of female academics, while countries with German university traditions (e.g., Sweden, Hungary, Germany, Finland, Austria, and Switzerland) had particularly low shares. We also document cross-country differences in female shares among full professors. A very similar ranking of countries emerges when we analyze female shares among full professors. The United States again stand out as the country with the highest share of female full professors. More broadly, Anglo-Saxon countries had the highest shares of female professors, while countries with German university traditions had particularly low shares. To the best of our knowledge, this is the first paper to uncover these facts, because similar data on academics has not been available up to this point.

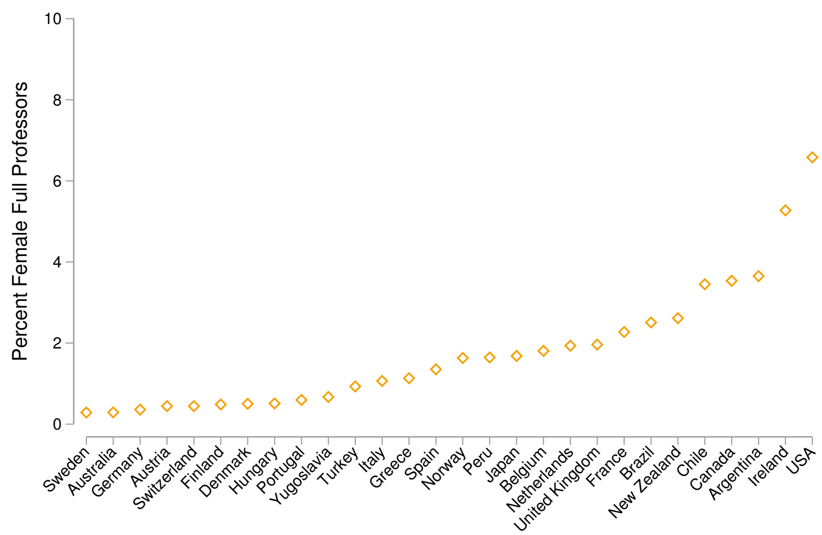
[23]For this figure, we calculate female shares at the cohort by country-level, e.g., United States in 1900 or United States in 1914, and then average these shares over the six cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort).

Figure 4: Percent of Female Academics by Country

(a) All Ranks



(b) Percent of Female Full Professors



Notes: The Figure shows the percentage of female academics by country. Panel (a) shows academics of all ranks. Panel (b) shows full professors, only. We calculate percentages of female academics at the cohort and country level, e.g. United States in 1900 or United States in 1914, and then average the percentages over the six cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort). The data were collected by the authors from various volumes of Minerva, see section 2 for details.

Hiring Gaps Across Selected Universities and at the City-Level

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 richness of the data, which also include many other excellent universities around the world. To select universities for this exercise, we rely on the well-known *Shanghai Ranking* of universities (see Ranking, 2020 for details). We choose the highest ranked universities in each country and report female shares for ten universities from the United States; five universities each from Germany and the United Kingdom, three universities each from Canada, France, Switzerland, and Japan; two universities from Italy, and one university each from Denmark, Australia, Sweden, Netherlands, Norway, Belgium, Ireland, Finland, Israel, Brazil, Argentina, Chile, Egypt, South Africa, Mexico, Poland, Portugal, and Spain.[24]

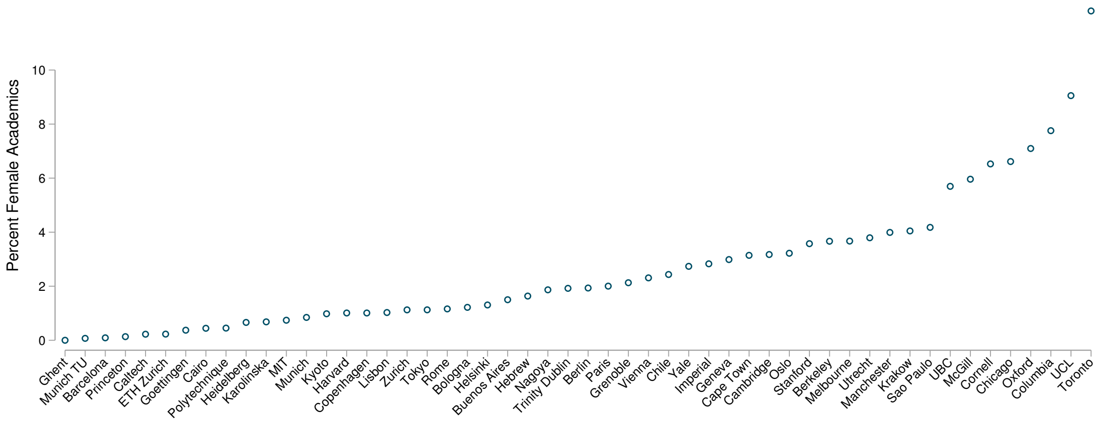
Interestingly, we observe similar patterns to those from the country-level analysis also among these highly ranked universities. On average, institutions in Anglo-Saxon countries had higher female shares than institutions in countries with German university traditions. Within the United States, we observe the highest female shares at Columbia, Chicago, and Cornell, while Harvard and MIT had comparatively low female shares. This closely aligns with historical accounts on female scientists in the United States (e.g., Rossiter, 1982).

We also analyze the evolution of female shares at the city-level. The leading role of the United States in hiring women is again clearly visible (Figure 6).

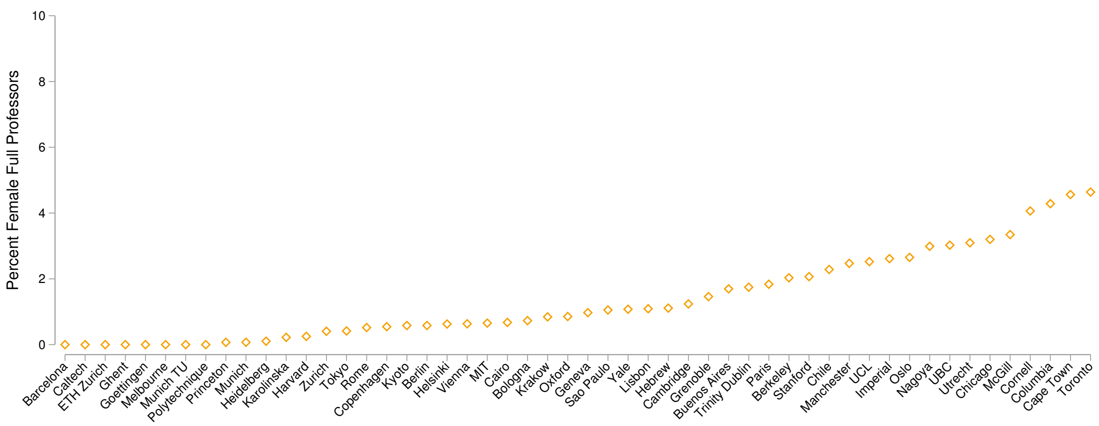
[24]The Shanghai Ranking ranks universities as of 2020. In many countries, e.g. the United States, the ranking of universities has remained very stable since 1900. In other countries the ranking has changed substantially. To report the most important institutions over the 20th century, we deviate somewhat from the Shanghai ranking for Germany and France. For Germany, we include the *University of Berlin* (now *Humboldt University*) instead of the *University of Bonn*, because the *University of Berlin* was the premier institution in Germany until WWII. In France, a number of reorganizations of universities occurred during the 20th century. To capture some of the highest ranked institutions during the course of the 20th century, we plot female shares for the *University of Paris*, *École Polytechnique*, and *Université de Grenoble*.

Figure 5: Percent of Female Academics by University

(a) All Ranks



(b) Percent of Female Full Professors

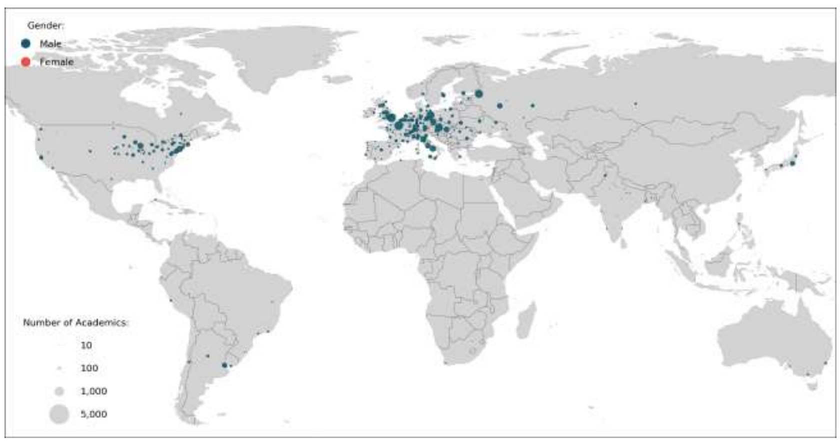


Notes: The Figure shows the percentage of female academics by universities. Universities were selected as explained in the text. Panel (a) shows academics of all ranks. Panel (b) shows full professors, only. We calculate percentages of female academics at the cohort and university-level, e.g. Harvard in 1900 or Harvard in 1914, and then average the percentages over the six cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort).

Figure 6: Number of Academics by Gender and City over Time (1900-1925)

(a) 1900

(b) 1914



(c) 1925

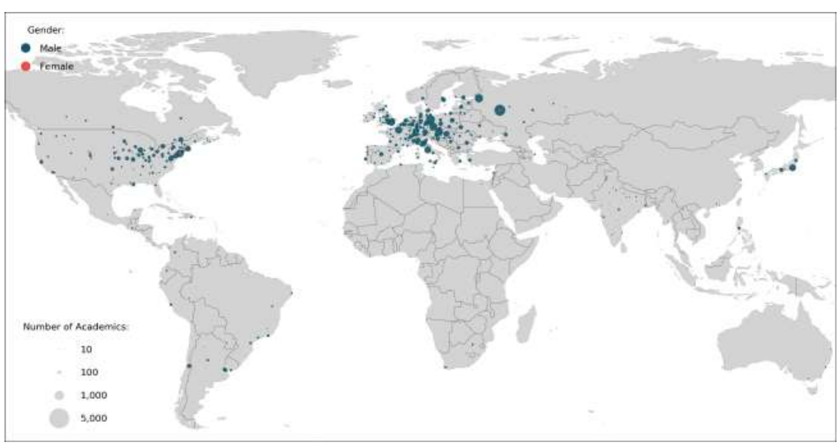
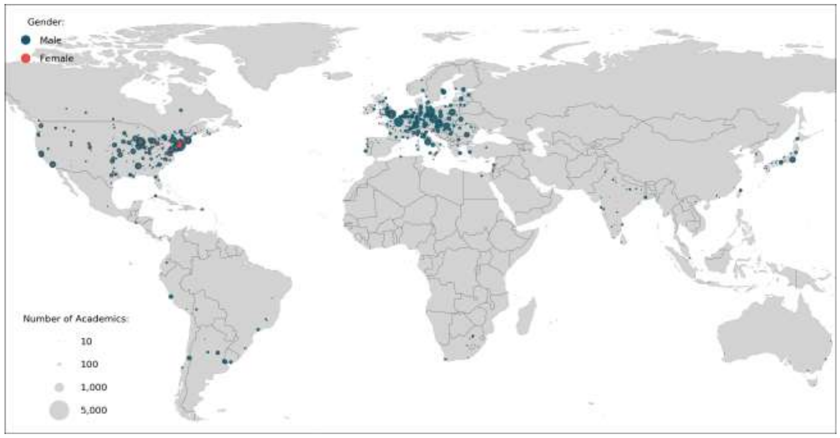
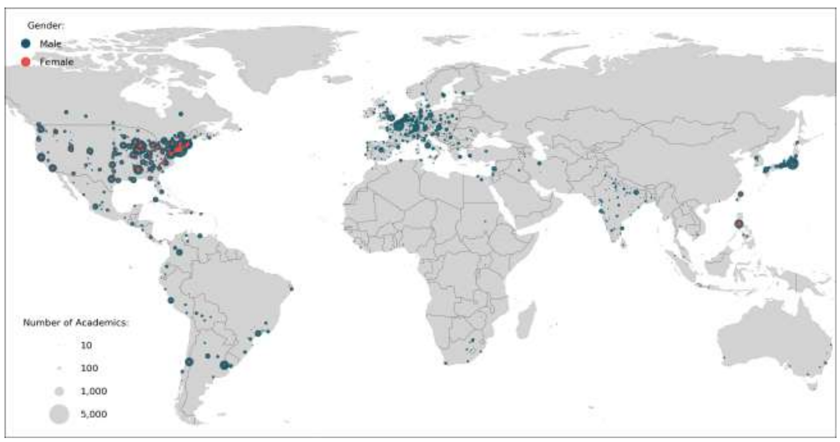


Figure 6: Number of Academics by Gender and City over Time (1938-1969)

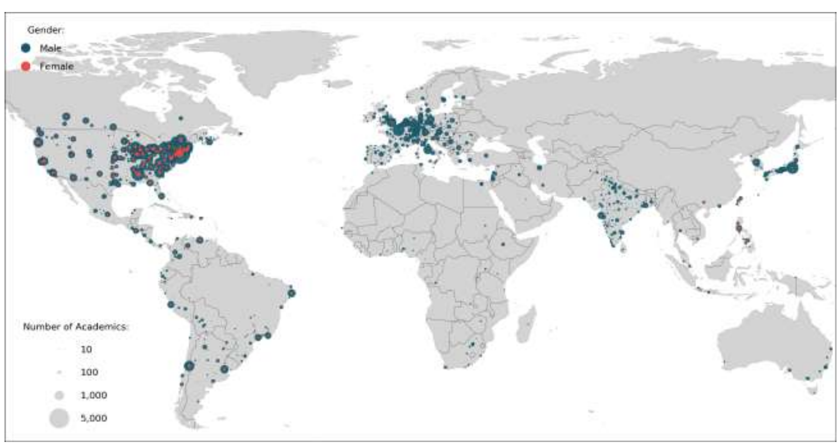
(d) 1938



(e) 1956



(f) 1969



Notes: The Figure shows the number of academics by gender and city, and its evolution from 1900 until 1969. The data were collected by the authors from various volumes of Minerva, see section 2 for details.

Hiring Gaps Across Disciplines

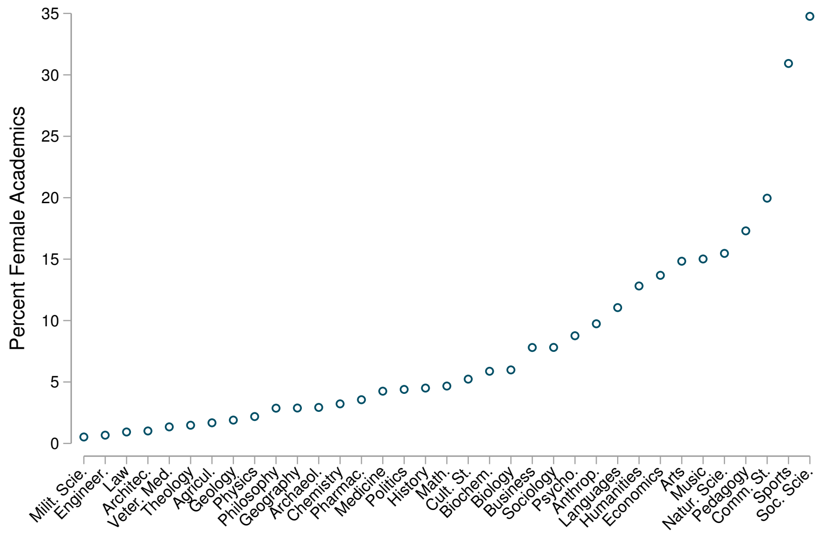
Our data also allow us to document differences in hiring gaps across disciplines (Figure 7).[25] Averaged over the period 1900 to 1969, no discipline had a female share greater than 35 percent. Most disciplines had female shares below 20 percent. Disciplines with particularly high female shares were natural sciences, pedagogy, communication studies, sports sciences, and social sciences.[26] Most of these were not research-oriented but rather focused on teaching undergraduates. Economics has a relatively high share of female academics, because it includes "Home Economics," a female dominated sub-discipline. The lowest female shares were in military sciences, engineering, law, architecture, veterinary medicine, and theology.

[25]For this figure, we calculate female shares at the cohort and discipline-level, e.g., physics in 1900 or physics in 1914, and then average these shares over the six cohorts (so that each cohort gets the same weight, independently of the total number of academics in that cohort).

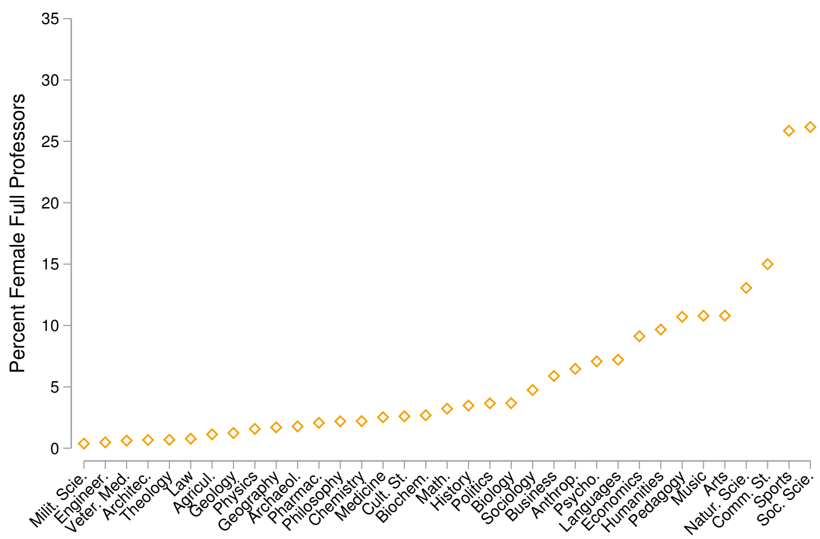
[26]Academics are only classified as natural scientists if they report their specialization as natural science, or similar. Academics who report a specialized discipline, e.g., physics or chemistry, are not classified as natural scientists but as physicists or chemists. Similarly, academics are only classified as social scientists if they do not report a specialized discipline.

Figure 7: Percent of Female Academics by Discipline

(a) All Ranks



(b) Full Professors



Notes: The Figure shows the percentage of female academics by discipline. Panel (a) shows academics of all ranks. Panel (b) shows full professors, only. We calculate percentages of female academics at the cohort and discipline-level, e.g. physics in 1900 or physics in 1914, and then average the percentages over the six cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort). The data were collected by the authors from various volumes of Minerva, see section 2 for details.

4 Gender Gaps in Publications

In this section, we explore gender gaps in publications. One of the unique advantages of studying academics is that we observe individual-level measures of output (publications) that are comparable across time and space. Publications are key metrics of performance that are commonly used, for example, to evaluate career progression, allocate research funds, and rank universities. Comparable performance measures are harder to come by in other professions. As previously discussed, we do not interpret publications as a measure of true ability of an academic. They reflect realized output that could be affected by preferences, discrimination, or other biases.

To estimate gender gaps in publications, we focus on the scientist sample, the sample of academics working in mathematics, physics, chemistry, biochemistry, biology, and medicine. We estimate the following "Mincer-type" regression:

$$\text{Pub}_{idt} = \beta_1 + \beta_2 \text{Female}_{idt} + \text{Fixed Effects} + \varepsilon_{idt}, \qquad (1)$$

where Pub_{idt} measures the number of papers that scientist i in cohort t and department d (e.g., physics at Harvard, which also determines i's country and discipline) published in the journals covered by the *Web of Science*.[27] As described above, we measure scientist i's papers in a $\pm$ five-year-window around i's cohort $t(i)$. I.e., for scientists that we observe in 1925, we consider papers published between 1920 and 1930.[28] In Appendix Table C.2, we repeat the analysis by measuring papers in a $\pm$ three-year-window around i's cohort $t(i)$. I.e., for scientists that we observe in 1925, we consider papers published between 1922 and 1928.[29] The main regressor of interest is the indicator variable Female_{idt}. All regressions control for a large number of fixed effects. In the baseline specification, we control for cohort, discipline, and country fixed effects. 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, and/or countries. In additional specifications, we control for the three-way interaction of these fixed effects. We also report specifications with university or department, or even for department $\times$ cohort fixed effects. It is important to note, that these regressions are only suggestive, because the university or

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

[28] The publication measure counts publications in the relatively high quality journals covered by the *Clarivate Web of Science*.

[29] Results show very similar gender gaps in publications. Note, the point estimates are lower because the mean number of publications is lower in a $\pm$ three-year-window than in a $\pm$ five-year-window.

department of an academic may be endogenous.[30] To account for the potential correlation of the residual ε_{idt} within country-discipline cells, e.g., chemistry in the United States, we cluster the standard errors at the country-discipline level.

Table 2: Individual-Level Publication Gaps

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable	Publications	Publications	Publications	Publications	Publications	Standard. Publications	Standard. Publications	Standard. Publications	Standard. Publications	Standard. Publications
Panel A: Full Set of Universities										
Female	-2.345***	-2.568***	-2.152***	-2.223***	-2.170***	-0.219***	-0.224***	-0.178***	-0.188***	-0.185***
	(0.414)	(0.468)	(0.486)	(0.475)	(0.484)	(0.012)	(0.012)	(0.011)	(0.013)	(0.017)
Observations	132,677	132,677	132,677	132,677	132,677	132,677	132,677	132,677	132,677	132,677
R-squared	0.087	0.127	0.170	0.201	0.234	0.003	0.003	0.062	0.093	0.140
Panel B: Stable Set of Universities										
Female	-3.107***	-3.371***	-3.168***	-3.180***	-3.062***	-0.313***	-0.303***	-0.279***	-0.274***	-0.263***
	(0.342)	(0.347)	(0.362)	(0.330)	(0.351)	(0.031)	(0.030)	(0.018)	(0.018)	(0.014)
Observations	69,147	69,147	69,147	69,147	69,147	69,147	69,147	69,147	69,147	69,147
R-squared	0.105	0.141	0.166	0.187	0.224	0.010	0.016	0.047	0.066	0.117
Cohort FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country FE	Yes					Yes				
Cohort × Discipline × Country FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist-level. Panel A reports results for scientists from all universities, while panel B reports results for scientists from universities observed in all six cohorts. In columns 1-5, the dependent variable equals the number of publications in a ±5 - year window around a Minerva cohort (i.e. 1909-1919 for a scientist listed in 1914). In columns 6-10, the dependent variable equals the number of publications, but standarized at the country-cohort-discipline level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman. The regressions also control for various fixed effects, as indicated in the table. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

The estimates indicate that women publish around 2.3 fewer papers compared to men (Table 2, panel A, column 1, significant at the 1 percent level). This is a substantial gap compared to the mean of publications, which is around 4.3.[31] The gap is robust to the inclusion of more restrictive fixed effects, which control for any difference at the cohort × country × discipline level.

Even comparing women to men within the same university or department (panel A, columns 3 and 4), the publication gap only shrinks marginally. Finally, in panel A, column 5, we control for cohort × department fixed effects. We thus estimate publication gaps for scientists in the same department and cohort, e.g., Harvard physicists in 1969. Even within this restricted group of scientists, women publish around 2 fewer papers than men.

[30] These controls serve a similar purpose as controlling for industry or occupation in standard Mincer regressions.

[31] Note that many women entered the data in later periods and work in the United States. In the later periods and in the United States the average scientist publishes more papers. We therefore also report results where we standardize publications at the country-cohort-discipline level (columns 6-10).

To provide a more intuitive interpretation of the publication gaps, we normalize the number of publications to have mean 0 and standard deviation 1 within each country, cohort, and discipline (e.g., physics in the United States in 1969). The results indicate a negative publication gap for female academics between 0.22 and 0.17 of a standard deviation (Table 2, panel A, columns 6-10, all significant at the 1 percent level).

We also show results for a sample of universities that we observe in all six cohorts to rule out that compositional changes drive our findings. In this sample, we estimate publication gaps that are at least as large as in the unrestricted sample of universities (Table 2, panel B). We also show that these findings are robust in a sample of scientists with unique lastname - first initial - discipline combination in every cohort (Appendix Table C.1).

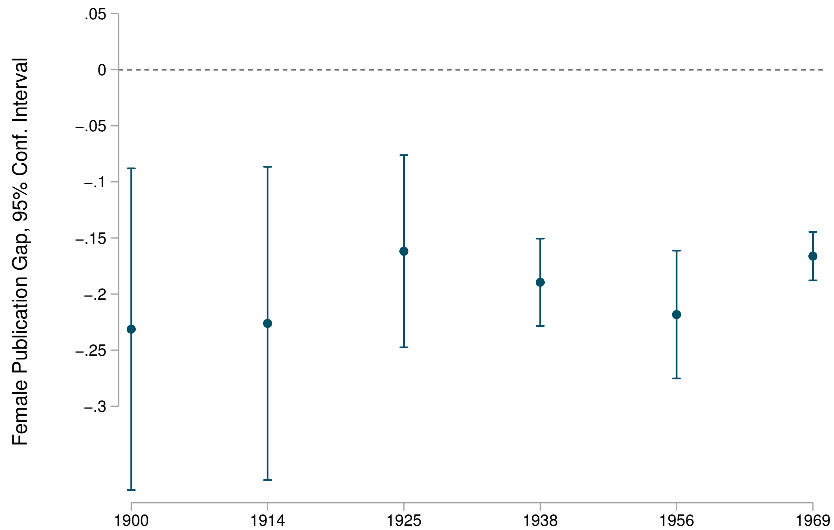
4.1 Publications Gaps Over Time

Next, we study the evolution of gender gaps in publications over time by interacting the female indicator with indicators for each cohort t:

$$\text{Pub}_{idt} = \beta_1 + \sum_{\tau=1900}^{1969} \beta_\tau \text{Female}_{idt} \times 1[t(i) = \tau] + \text{Fixed Effects} + \varepsilon_{idt}. \tag{2}$$

We then plot the six coefficients $\hat{\beta}_\tau$ and corresponding 95 percent confidence intervals in Figure 8. Confidence intervals are relatively large for the early cohorts due to the relatively small number of female scientists in these cohorts. The publication gaps are relatively stable over time and hover around 0.17 of a standard deviation, suggesting that gender gaps in publications did not narrow over the first 70 years of the 20th century.

Figure 8: Individual-Level Publication Gaps over Time



Notes: The Figure shows gender gaps in publications over time. The gender gaps are estimated with equation 2. We then plot $\hat{\beta}_\tau$ and the corresponding 95 percent confidence interval for each of the six cohorts: 1900, 1914, and so on. The regressions are estimated with country cohort discipline and department fixed effects. The lack of a trend in publication gaps is robust to different specifications of the fixed effects.

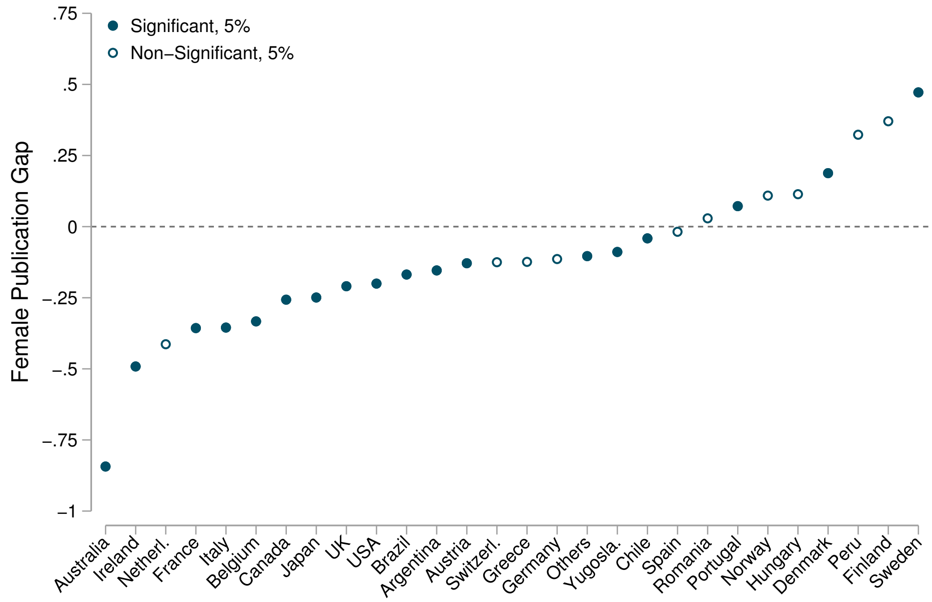
4.2 Publication Gaps by Country

In further results, we estimate gender gaps in publications by country:

$$\text{Pub}_{idt} = \beta_1 + \sum_{\omega} \beta_\omega \text{Female}_{idt} \times 1[\text{Country}(i) = \omega] + \text{Fixed Effects} + \varepsilon_{idt}. \tag{3}$$

To ease comparisons, we report results that use the standardized publication measures. In most countries, the publication gaps are negative. However, there are also a few countries with positive publication gaps, especially those with German university traditions (Figure 9).

Figure 9: Individual-Level Publication Gaps across Countries



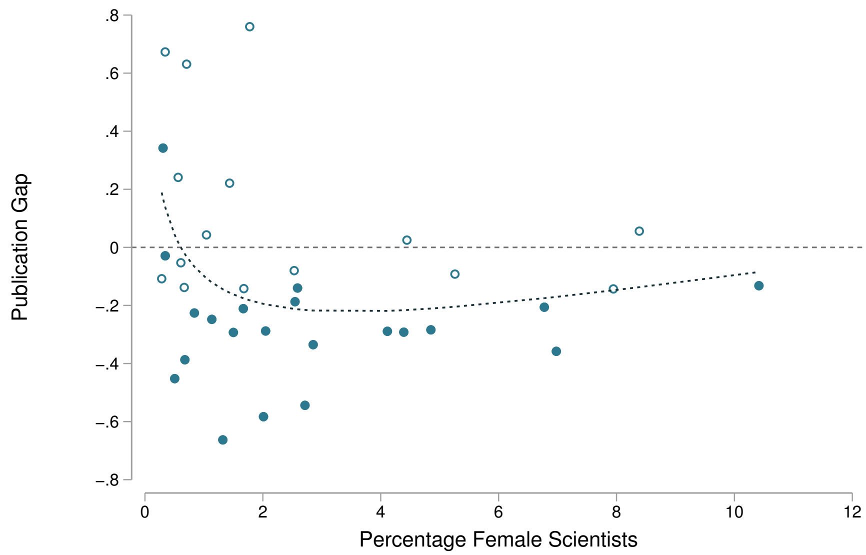
Notes: The Figure shows gender gaps in publications across countries. The gender gaps are estimated with equation 3 including cohort × country × discipline and department fixed effects. We then plot the estimated $\hat{\beta}_\tau$ for each country. Filled dots correspond to estimates that are significant at the 5 percent level. Circles correspond to estimates that are not significant at the 5 percent level. We estimate separate $\hat{\beta}_\tau$'s for countries with at least 10 women over all six cohorts, plus one $\hat{\beta}_\tau$ for all other countries combined, which is not reported in the Figure. We show the results with country cohort discipline and department fixed effects. The ranking of countries remains very similar if we used different sets of fixed effects.

4.3 Publication Gaps And The Share of Female Academics: "Gender Swoosh"

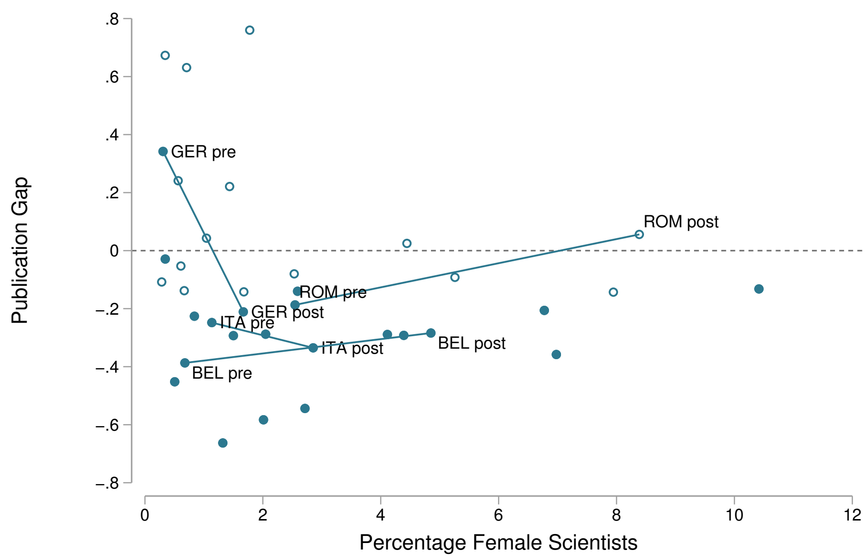
Cross-country variation in gender gaps in publications and in hiring of female academics allow us to analyze the relationship between publication gaps and the share of female scientists in a country. For this analysis, we estimate country-specific publication gaps separately for two time periods: pre-WWII (cohorts 1900, 1914, 1925, and 1938) and post WWII (cohorts 1956 and 1969). We then relate the estimated publication gaps to the percentage of female scientists in each country and period.

Figure 10: Individual-Level Publication Gaps by Country and Time Period and the Share of Females

(a) Non-Linear Fit



(b) Highlighting Selected Countries



Notes: The Figure relates estimated gender gaps in publications to the percentage of female scientists by country and time period. Each dot corresponds to a country and one of two periods: pre-WWII and post-WWII, e.g. United States - Pre WWII or United States - Post WWII. Gender gaps in publications are estimated by country with the equivalent of equation (3) but estimating two coefficients (one for the pre-WWII period, one for the post-WWII period) per country. Countries with at least two women per period are plotted in the Figure. Filled dots correspond to an estimate of a gender gap that is significant at the five percent level. Circles correspond to an estimate of a gender gap that is not significant at the five percent level. Panel (a) also shows a non-linear regression line estimated by fractional polynomials. Panel (b) shows examples of countries.

Figure 10, panel (a), plots the estimated publication gaps as a function of the percentage of female scientists in each country and period. The figure suggests a "gender swoosh" pattern (i.e., the Nike logo). Publication gaps are mostly positive in countries and periods with very low shares of female scientists.[32] This can be referred to as the "Marie-Curie" period: only exceptional women were hired and, on average, they published more than men despite potential discrimination in the publication market. With increasing shares of women in the profession, gender gaps in publications turned negative. However, when the share of women increased beyond very low levels, the negative gender gaps in publications narrowed. This suggestive relationship holds within many countries (Figure 10).

In Appendix 10.2, we outline a model along the lines of Roy (1951) to interpret the relationship between the share of women in academia and the observed gender gap in publications, i.e. the "gender swoosh" (Figure 10a). The proposed model allows for (i) selection on unobservables in the hiring market, (ii) gender bias in hiring, and (iii) gender bias in the publication market. These features make scientists' publications a function of the share of women in a country-period because of (a) *indirect* effects of selection and gender bias in the hiring market and (b) the *direct* effects of gender bias in the publication market. Based on the model, we specify a semiparametric regression that directly estimates the *total* gender gap in publications (indirect + direct effects) as a function of the share of female scientists in a country-period. Estimates of this regression (Appendix Figure C.1) illustrate a qualitatively similar pattern to that in Figure 10a. Publication gaps are smallest in countries and periods with very low shares of female scientists. With increasing shares of women in the profession, gender gaps in publications become more negative. However, when the share of women increases beyond extremely low levels, the negative gender gap in publications narrows. These results indicate that a standard Roy model can approximate the pattern of the "gender swoosh" and that gender biases in hiring may have indirect repercussions on the observed productivity of female scientists.

5 Citation Gaps: Controlling for Predicted Citations

In the previous section, we have shown evidence that women published fewer papers, even conditional on cohorts, countries, disciplines, universities, or departments. In this section, we

[32]In countries and periods with very low female shares the data contain very few (often less than five) women and, hence, most of the estimates are not significantly different from zero at conventional levels. Nonetheless, in the group of countries and periods with very low female shares there are three coefficients with relatively low p-values: Germany: coefficient 0.341, p-value 0.002, Sweden: coefficient 0.241, p-value 0.137, and Austria: coefficient 0.673, p-value 0.166.

explore if papers written by women were cited less. For this analysis, we turn to a paper-level analysis using all papers that we merged to the scientists in our data.[33]

5.1 A Novel Procedure to Predict Citations

A key ingredient of our analysis of citations gaps is a newly developed supervised machine learning model which creates a new measure of predicted citations. We then control for the citation potential of a paper using the predicted citations measure.[34] We train a model that uses the words in the titles of papers to predict citations.[35] The model learns complex relationships between research topics and citations (see Appendix 11 for further details on the model).

In preparation for the machine learning step, we clean all non-alphanumerical characters from the papers' titles and remove very common words (stopwords, e.g. "and"). Next, we extract all unigrams (i.e. words) and bigrams (i.e. two-word combinations) from the title of each paper to obtain a paper-1,2-gram-matrix $\boldsymbol{X}$ with entries x_{pj}, where p denotes papers and j denotes unigrams and bigrams. As is common in text-based machine learning, we then reweight the matrix using term-frequency inverse-document frequency (tf-idf) reweighting. The reweighting 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 optimizes the following objective function:

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

where y_p are the total citations of paper p, which we transform into permilles (1000 quantiles) within each country, discipline, and cohort.[36] The main explanatory variables are the unigrams and bigrams that correspond to the respective entries of the paper-1,2-gram-matrix $\boldsymbol{X}$. We additionally include a full set of indicators for the number of words in the title of each

[33] An analysis of citations at the individual level would conflate two types of gender gaps: 1) gender gaps in publications (as documented in the previous set of results) and 2) gender gaps in citations to each published paper which we explore in this section.

[34] We make a pre-trained model of predicted citations available at carloschwarz.eu/programming/. The model allows to predict the log number of citations from the titles of papers. We also provide a Python and Stata wrapper.

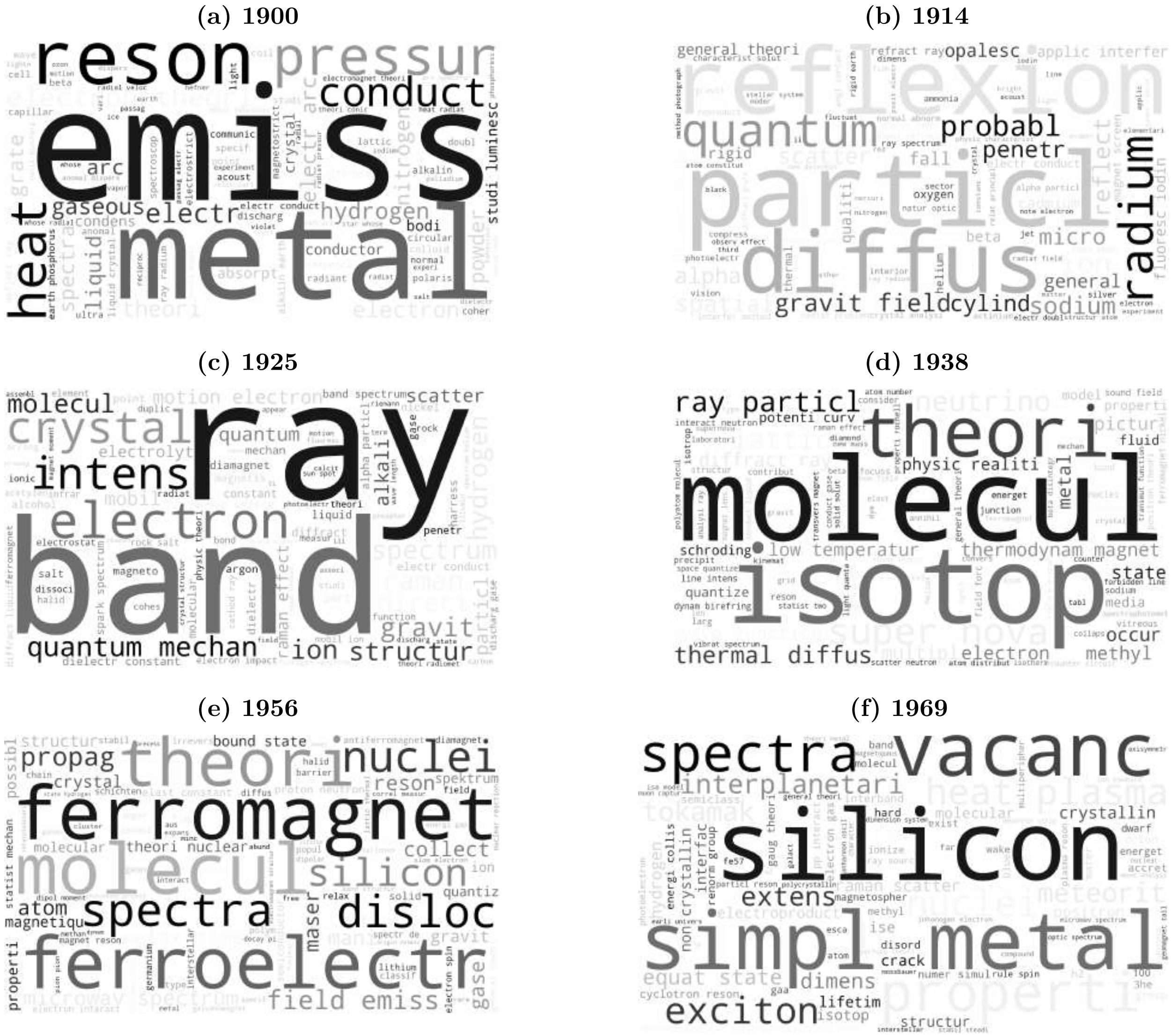
[35] Importantly, the *Web of Science* translates almost all titles into English.

[36] Because of outliers in citations, we transform p's total number of citations into permilles of the citation distribution within a discipline, country, and cohort (e.g. physics in the United States in 1900). If we instead use citation counts, the estimated gender gap in citations remains very similar. See Appendix (11) for details, as well as additional robustness checks on the predicted citations procedure.

paper in the model. To prevent the model from internalizing biases against papers published by women (see Barocas and Selbst (2016) for an overview of AI biases), the training sample solely consists of papers published by men. The method therefore allows to predict the citations of each paper, as if it had been written by men. We train separate models for each discipline in each of the six cohorts. For each discipline and cohort, we choose the optimal normalization strength λ using 10-fold cross validation. The algorithm predicts citations $\hat{y}_p$ for each paper p.

The model identifies highly intuitive relationships between words and citations. Figure 11 summarizes the unigrams and bigrams that predict high citations for physics and how they changed between 1900 and 1974. For example, for the 1909-1919 period, words like “particle”, “quantum”, and “re-flexion“ predict high citations. All of these terms relate to the study of light waves and the photoelectric effect for which Einstein received the Nobel Prize in 1921. Appendix 11.1 shows unigrams and bigrams that predict high citations for other disciplines. An example of a highly cited word in biology is “chromosom” (1900) which was crucial in the theory of inheritance which was developed in 1902 by Walter Sutton and Theodor Boveri. Another example is “radioimmunoassay” (1969) which describes the diagnostic method which earned Rosalyn Sussman Yalow the physiology/medicine Nobel prize in 1977. Examples for chemistry are the words “ketene” (1914/25), which were first discovered by Hermann Staudinger in 1905, “polarograph” (1938) which was invented by Jaroslav Heyrovský, and “macrocycles” (1969) which were discovered by Charles J. Pedersen in 1967. All of these discoveries eventually let to Nobel prizes in chemistry (Staudinger 1953, Heyrovský 1959, Pedersen 1987). An example for mathematics is the bigram “dirichlet principle” (1900), which was popularized by Hilbert in 1904 after he developed a direct method in the calculus of variations.

Figure 11: Words that Predict High Citations in Physics over Time



Notes: The Figure shows the unigrams and bigrams that predict the highest citations in physics for each of the Minerva 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 5.1 for details.

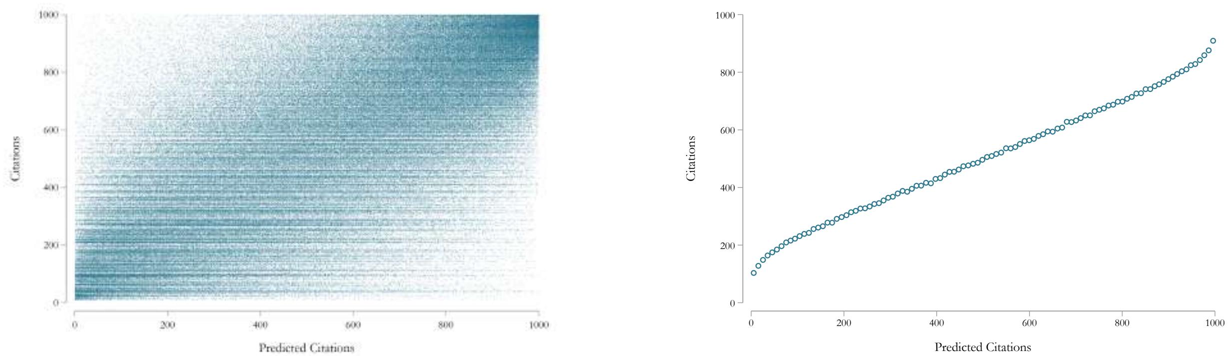
In addition to identifying intuitive relationships between words and citations the model performs well in predicting a paper's actual citations. Figure 12 visualizes the strong positive relationship between predicted and actual citations (R^2 =0.43).[37] Of course, even papers with titles that use word combinations that are often highly-cited (e.g., quantum) sometimes receive no citations at all. Similarly, papers with titles that use word combinations that mostly receive very few citations, sometimes receive a very high number of citations. Hence, the scatter plot shows a large number of points off the diagonal (panel (a)). On average,

[37]Note that this is the within sample R^2. The method also performs well if we use an "out of sample" approach (see section 5.2).

however, papers with few predicted citations also receive few actual citations as we visualize using a binscatter plot (panel (b)).

Figure 12: Predicted and Actual Citations

(a) Raw Data **(b) Binned Scatter Plot**



Notes: The Figure shows the relationship between actual and predicted ciations in permilles. Actual citations are the citation permille of each paper in the data. Permilles are calculated at the cohort-country-discipline level. Permilles of predicted citations are estimated with an L2-regularized regression model (ridge regression) that uses unigrams and bigrams of the title as inputs, see section 5.1 for details. Panel (a) reports the raw data, i.e. each dot represents one of the more than half a million papers. Panel (b) reports a binned scatter plot, with 100 bins (i.e., 10 permilles per bin).

5.2 Paper-Level Citation Gaps — Controlling for Predicted Citations

In the following analysis, we estimate the citation gaps at the paper-level, depending on whether papers were published by men or women. Importantly, we control for predicted citations to control for the fact that women may be working on different topics than men.[38] We estimate the following regression:

$$\text{Citation Permille}_p = \gamma_1 + \gamma_2 \text{Female}_p + \sum_{k=1}^{1000} \gamma_k \hat{y}_p^k + \text{Fixed Effects} + \xi_p. \tag{5}$$

The dependent variable is the citation permille of paper p. The main explanatory variable is Female_p, the paper-level share of authors who are female.

For most papers Female_p is either 0 or 1 (see Appendix Figure D.1). Importantly, we control for the predicted citations of paper p by including an indicator $\hat{y}_p^k$ for each permille k of

[38] In a recent paper Koffi (2021) proposes a different methodology to estimate whether papers by female authors are under-cited. Her methodology uses text similarity to identifying existing papers that the focal paper should have cited. In contrast, our methodology controls for differences in citation potential of papers. A key advantage of our methodology is that it can be used even if female shares are very low.

the predicted citation distribution. All regressions additionally contain different sets of fixed effects.[39]

Citations to papers by female authors receive on average 27 permilles (i.e., 2.8 percentiles) fewer citations than papers by male authors (Table 3, column 1, significant at the 1 percent level). Adding additional fixed effects reduces the estimated citation gap somewhat, but even controlling for department×cohort fixed effects, papers by female authors receive on average 22 permilles fewer citations. Flexibly controlling for predicted citation permilles reduces the citation gap by around 15 percent. The gap remains, however, highly statistically significant (Table 3, columns 6-10, all significant at the 1 percent level).

The results indicate that papers by female authors receive fewer citations, even if one controls flexibly for the types of papers they write. This suggests that women do not receive fewer citations because of a different choice of topics compared to men.
In appendix Table D.2, Panel A, we show that our findings are robust to using citation counts (instead of permilles) as the outcome.[40] Further, as we can only assign the gender to authors listed in *Minerva*. We also show that our results are very similar if we restrict the sample to papers for which we know the gender of all authors (Table D.2 Panel B). In the full set of papers we only know the gender of authors who are part of *Minerva* but not of other authors (e.g., graduate students). This measurement error in the explanatory variable, would most likely lead us to underestimate the effect of gender on citations. Another concern could be that the model absorbs more variation for male authored papers, as the model is trained for male authors. We address this concern by training an "out of sample" version of our model. This model is trained on a 20% holdout sample for each cohort and subject. This approach comes at the cost of excluding papers from the outcome regression. Citation gaps using the "out of sample" model are slightly larger in absolute magnitude than for our standard model (Table D.2 Panel C).

[39] For these results the country, university and department fixed effects are defined at the paper-level, i.e., a paper coauthored by scientists from Harvard and MIT has a separate fixed effect from papers authored by Harvard scientists, only.

[40] As usual when working with citation data, a significant fraction of papers does not receive any citation (12% on average in our case). We show that this does not affect the results (Appendix 11.2).

Table 3: Paper-Level Citations Gaps: Controlling for Predicted Citations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Without Predicted Citations					With Predicted Citations				
Dependent Variable	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles
Share of Female Authors	-27.180***	-27.866***	-17.783***	-22.095***	-22.932***	-19.204***	-19.117***	-18.567***	-18.841***	-19.376***
	(4.035)	(4.096)	(5.362)	(3.802)	(4.237)	(3.064)	(3.140)	(2.540)	(2.603)	(2.569)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830
R^2	0.002	0.000	0.039	0.052	0.071	0.508	0.509	0.518	0.520	0.523

Notes: The Table shows gender gaps in citations per paper. Results are estimated at the paper-level. The dependent variable is equal to the permille in the cohort-country-discipline citation distribution of a paper. The main explanatory variable is the share of female authors of the paper. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for 1000 indicators that equal 1 if the paper falls into a certain permille of the predicted citation distribution. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model (ridge regression), see section 5.1 for details. Standard errors are computed by boostrap methods and clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

5.3 Explanations for Paper-Level Citation Gap

In additional results, we explore further reasons that could explain why papers by female authors may receive fewer citations. After controlling for the detailed topics of papers, there are at least three additional reasons why papers by female scientists may receive fewer citations. First, women may have fewer opportunities to write papers with co-authors (production effect). Such a production effect may translate into fewer citations, because coauthored papers on average receive more citations (e.g., Wuchty, Jones, and Uzzi, 2007). Second, women may publish their papers in lower ranked journals because of biased editors or referees (publication effect). This has been shown for economics papers (Card et al., 2020b). A third possibility is that papers of female authors receive fewer citations because of biases in the citation market. Biases could arise, for example, because women have fewer opportunities to present their work or because of discrimination.

We explore the first explanation by controlling for the number of authors of each paper (i.e., a fixed effect if the paper has one author, another fixed effect if the paper has two authors, and so on). Interestingly, controlling for the number of authors does not affect the female citation gap (Table 4, column 2).

Next, we explore the publication effect by including a full set of journal fixed effects (Table 4, column 3). Of course, the journal is potentially endogenous and the estimates should therefore be interpreted as a decomposition of the citation gap. The inclusion of journal fixed effects reduces the female citation gap by about a third, suggesting that a third of the citation gap can be explained by the fact that papers by women are published in journals that receive somewhat fewer citations. Notably, even after controlling for the number of co-authors, the journal, and the research topic with our measure of predicted citations, papers by female authors still receive significantly fewer citations than papers by their male peers within the same department (Table 4, column 8, significant at the 1 percent level).

Table 4: Paper-Level Citations Gaps: Controlling for Number of Authors and Journals

Dependent Variable	(1) Citations Permilles	(2) Citations Permilles	(3) Citations Permilles	(4) Citations Permilles	(5) Citations Permilles	(6) Citations Permilles	(7) Citations Permilles	(8) Citations Permilles
Share of Female Authors	-22.932***	-22.793***	-13.023***	-13.639***	-19.376***	-19.476***	-12.728***	-12.962***
	(4.237)	(4.090)	(3.104)	(2.833)	(2.569)	(2.501)	(3.252)	(3.155)
Cohort × Department FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nr. Authors FE		Yes		Yes		Yes		Yes
Journal FE			Yes	Yes			Yes	Yes
Predicted Citations (1000 quantile) FE					Yes	Yes	Yes	Yes
Observations	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830
R^2	0.071	0.102	0.260	0.271	0.523	0.525	0.560	0.561

Notes: The Table shows gender gaps in citations per paper. Results are estimated at the paper-level. The dependent variable equals the permille in the cohort-country-discipline citation distribution of a paper. The main explanatory variable is the share of female authors of the paper. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for 1000 indicators that equal 1 if the paper falls into a certain permille of the predicted citation distribution. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model (ridge regression), see section 5.1 for details. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

6 Gender Gaps in Promotions

In the last section, we investigate gender gaps in academic promotions. For these results, we focus on the sample of academics who were not already full professors when they entered the dataset in cohort t.[41] We then analyze whether they get promoted to full professor by cohort $t+1$ (see Appendix 8.1.1 for more information on the coding of promotions). Promotions to full professor are particularly important because, in most countries, full professors have unique privileges and particularly high job security and salaries. We present results for academics in all disciplines and then for the scientist sample. In the scientist sample, we can control for fine-grained publication and citations records. We estimate the following regression:

$$\text{Promotion Full Prof}_{idt} = \pi_1 + \pi_2 \text{Female}_{idt} + \text{Fixed Effects} + \upsilon_{idt}. \qquad (6)$$

[41]This results in a smaller sample because academics who enter the data as full professors cannot be included in the analysis. Furthermore, all academics who enter the data in the last cohort (independently of their rank) cannot be included in the analysis. Lastly, we can only analyze promotions for academics who we observe in at least two cohorts of the data.

The dependent variable Promotion Full Prof_{idt} is an indicator for whether academic i who entered the data in department d and cohort t was promoted to full professor by cohort $t+1$ (in any department). Among all academics, women are about 22 percentage points less likely to be promoted to full professor than men (Table 5, panel A, column 1, significant at the 1 percent level). Because the probability of promotion is around 64 percent, women are about 34 percent less likely to be promoted. The large gender gap in the probability of promotion to full professor is robust to the inclusion of more stringent fixed effects. Even compared to men within the same department and cohort, women are around 20 percentage points (or 31 percent) less likely to be promoted (Table 5, panel A, column 4, significant at the 1 percent level).

In the scientist sample, women are 20 percentage points less likely to be promoted to full professor (Table 5, panel A, column 5, significant at the 1 percent level). Crucially, in this sample we can also control for the scientist's publication and citation record. While publications have a significant effect on promotions, controlling for these records hardly affects the gender gap in the promotion to full professor.[42] If we control for publications and citations, and compare men and women within the same department and cohort, women are still 18 percentage points less likely to be promoted to full professor (Table 5, panel A, column 9, significant at the 1 percent level). The unexplained gender gap in promotions to full professor is larger than the effect of a four standard deviations worse publication record. This is particularly striking because the true quality of women conditional on the same number of publications and citations should be, if anything, higher in the presence of discrimination and other biases in the publication and citation market.

As above, we repeat the analysis on the set of universities that we observe in all cohorts. Among all academics, the gender gap in the probability of promotion to full professor is very similar in the stable set of universities. Among the scientists, the probability of promotion to full professor is around 15 percentage points and highly significant in the stable set of universities(Table 5, panel B).

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

Table 5: Promotion Gaps: Full Professor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All Academics				Scientists				
Dependent Variable:	Promotion	Promotion	Promotion	Promotion	Promotion	Promotion	Promotion	Promotion	Promotion
Panel A: Full Set of Universities									
Female	-0.224***	-0.217***	-0.200***	-0.195***	-0.205***	-0.194***	-0.189***	-0.193***	-0.185***
	(0.022)	(0.024)	(0.024)	(0.028)	(0.042)	(0.047)	(0.054)	(0.065)	(0.065)
Std. Publications									0.045***
									(0.008)
Std. Citations									0.007
									(0.004)
Observations	31,829	31,829	31,829	31,829	12,822	12,822	12,822	12,822	12,822
R-squared	0.133	0.229	0.453	0.563	0.164	0.240	0.425	0.540	0.546
Panel B: Stable Set of Universities									
Female	-0.226***	-0.208***	-0.229***	-0.220***	-0.163***	-0.142***	-0.154***	-0.167***	-0.156***
	(0.022)	(0.023)	(0.032)	(0.033)	(0.032)	(0.033)	(0.044)	(0.048)	(0.045)
Std. Publications									0.053***
									(0.007)
Std. Citations									0.007
									(0.004)
Observations	18,580	18,580	18,580	18,580	8,273	8,273	8,273	8,273	8,273
R-squared	0.188	0.308	0.446	0.574	0.193	0.284	0.403	0.533	0.542
Cohort FE	Yes				Yes				
Discipline FE	Yes				Yes				
Country FE	Yes				Yes				
Cohort × Discipline × Country FE		Yes	Yes			Yes	Yes		
Department FE			Yes				Yes		
Cohort × Department FE				Yes				Yes	Yes

Notes: The Table shows gender gaps in the probability of promotion to full professor. Results are estimated at the academic-level. Panel A reports results for academics from all universities, while panel B reports results for academics from universities observed in all six cohorts. The dependent variable is an indicator that equals 1 if an academic who entered the dataset in cohort t was promoted to full professor by cohort $t+1$. The main explanatory variable is an indicator that equals 1 if the academic is a woman. Columns 1-4 report results for academics in all disciplines, while columns 5-9 report results for academics in the sciences. The regressions also control for various fixed effects, as indicated in the table. For the scientist sample, the regressions also control for the publication and citation record of the scientist. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

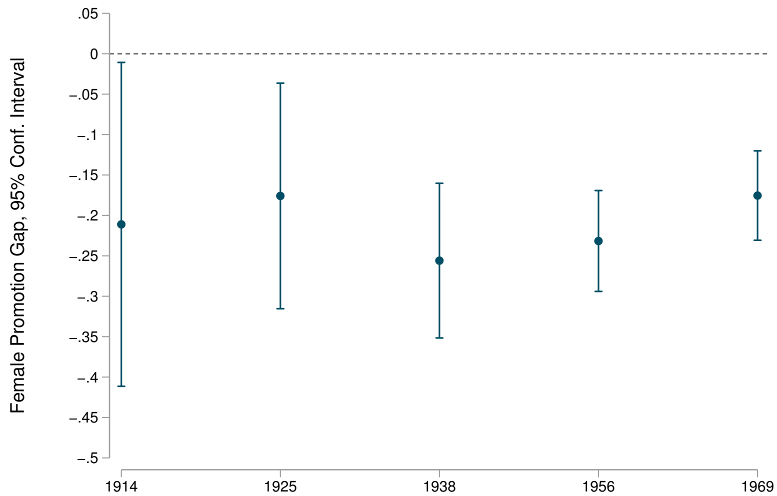
6.1 Promotion to Full Professor Over Time

We also investigate the evolution of the gender gap in promotions to full professor over time with the following regression:

$$\text{Promotion Full Prof}_{idt} = \pi_1 + \sum_{\tau=1914}^{1969} \pi_\tau \text{Female}_{idt} \times 1[t(i) = \tau] + \text{Fixed Effects} + \upsilon_{idt}. \quad (7)$$

As we evaluate academics who are promoted between cohort t and cohort $t+1$, we cannot estimate the probability of promotion in the first cohort. In all cohorts, women are less likely to be promoted to full professor (Figure 13). Naturally, the small numbers of female academics during the first decades of the data lead to relatively large confidence intervals for those periods. Between 1914 and 1938, the promotion gap hovers between 25 and 20 percentage points and then starts declining to about 17 percentage points by 1969.

Figure 13: Promotion Gaps over Time



Notes: The Figure shows gender gaps in promotion to full professor over time. The gender gaps are estimated with regression 7 on the sample of all academics. We plot $\hat{\pi}_\tau$ and the corresponding 95 percent confidence interval for each of the five cohorts after the first: 1914, 1925, and so on. As we evaluate academics who are promoted between cohort t and cohort $t+1$, we cannot estimate the probability of promotion in 1900.

7 Conclusion

Leveraging new worldwide data on academics, this paper sheds light on the evolution of gender gaps in academia. From our analysis, four results stand out. First, only one percent of academics were women in 1900 and the share of women increased to only 11 percent by 1969. Anglo-Saxon countries were more successful in hiring female academics compared to countries with a German university tradition. Second, we document large and persistent gender gaps in publications. Third, our analysis of citation gaps suggests that papers by female authors receive fewer citations. These citation gaps are not easily explained by differences in the research topics women are working on, the journals that the paper was published in, or the number of co-authors. Lastly, the fact that female academics were less likely to be promoted, even compared to male peers within the same department, cohort, and with same publication and citation record are indicative of the unequal opportunities that pervaded academia.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic*

Journal: Applied Economics 2 (3):228–55.

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

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

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

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

———. 2020b. "Gender Differences in Peer Recognition by Economists." .

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

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

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

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

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

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

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

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko. 2006. "The homecoming of American college women: The reversal of the college gender gap." *Journal of Economic perspectives* 20 (4):133–156.

Hengel, Erin. 2020. "Publishing while female (summary)." .

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

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

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

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

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

Moser, Petra and Scott Kim. 2021. "Women in Science. Lessons from the Baby Boom." .

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

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

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

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

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

Roussille, Nina. 2020. "The central role of the ask gap in gender pay inequality." .

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

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

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

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

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

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

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

Wigner, Eugene P. 1972. "Maria Goeppert Mayer." *Physics Today* 25 (5):77.

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

Appendix

The Appendix presents further details on data, the predicted citation measure, and additional results:

- Appendix 8 provides details on the data collection.
- Appendix 9 shows additional results on hiring gaps.
- Appendix 10 shows additional results on publication gaps.
- Appendix 11 provides additional details on the predicted citations model.

8 Further Details on Data

8.1 Data on the Universe of University Academics from Minerva

8.1.1 Additional Information on the Coding of Academic Ranks

Minerva reports academic ranks for most academics. The ranks are reported either in original language (e.g., maître de conférence) or are translated into English or German. Overall, the source reports almost 4,000 different combinations of country and rank. We recode them at the country-level, because certain labels of ranks do not necessarily describe the same academic rank across countries. E.g., a *lecturer* in the British system has a higher academic rank than a *lecturer* in the U.S. system. We classify all positions into the following categories: professorial admin position (e.g., dean or head of department), full professor, associate professor, assistant professor, honorary professor, clinical faculty, visiting professor, teaching position, Emerita/us, Emerita/us associate professor, Emerita/us assistant professor. In a few cases, the source jointly lists a number of academics who hold different academic ranks (e.g., associate and assistant professors) without distinguishing the exact rank of each academic. In these cases, we assign the highest listed rank to each academic.

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

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

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

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

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

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

Promotion to full professor. We classify academics who enter the data at ranks 2, 3, or 4 in cohort t and are promoted to rank 1 as promoted to full professor.

8.1.2 Additional Information on the Coding of Disciplines

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

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

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

8.1.3 Identifying Academics with Multiple Appointments within a City

We identify academics with multiple appointments within a city by hand-checking all academics with duplicate surnames within a city. We then decide whether two Minerva entries indeed refer to the same academic based on the first name, specialization, academic rank and title. In cases, in which the information indicate that an academic indeed holds two appointments we harmonize, if necessary, the first name and collapse the two Minerva entries into a single observation. The resulting observation then contains the information on all appointments and specializations of an academic within a city.[43]

[43] In very rare cases, academics hold multiple appointments in different cities or even countries in the same *Minerva* cohort... As verifying that these observations are indeed only one academic cannot be done with certainty, we treat them as two separate observations. We show that all results are very similar in a sample

8.1.4 Linking Academics Across Minerva Cohorts

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

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

1. Universities sometimes report first names with slight variations across Minerva cohorts. E.g., the University of Leipzig reported the geographer Joseph Partsch as Jose*ph* Partsch in the 1914 Minerva but as Jose*f* Partsch in the 1925 Minerva.

2. In certain Minerva cohorts, some universities only report their professors using an abbreviated first name plus the surname. In other Minerva 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 Minerva but as *Johannes* Witte in the 1928 Minerva.

3. In certain Minerva cohorts, some universities only report their professors using initials plus the surname. In other Minerva 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 Minerva 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.

Within departments In this part, we explain how we link academics who remain in the same department between cohort t and cohort $t+1$. In a first step, we consider academics who stay in the same department between two Minerva cohorts. We link all academics from discipline d, country c, university u, and Minerva cohort t to all academics from the same discipline d, same country c, same university u, and cohort $t+1$ based on the academic's surname, the first initial, her discipline (e.g., physics, geography, and so on), and her university (which implicitly also matches on the country). Hence, all potential links we

of individuals with unique combinations of surname, first initial, and discipline. This indicates that such duplicates have only limited effect on the results.

consider have identical surnames, disciplines, and universities (and countries), but the first name is not necessarily identical. We then link academics across Minerva cohorts as follows:

1. If the entire information on the first name (in some cases the information on the first name that Minerva reports for that particular academic may be one or more initials in both Minerva cohorts) is identical in both Minerva cohorts, we consider these academics as linked across the two Minerva cohorts.

If the information on the first name differs across the two Minerva cohorts, research assistants examine each potential link and decide whether the academics are linked. E.g., the data contain the following data points for the Minerva cohorts 1925 and 1938:

Table A.1: Examples Within Department Merge

	Minerva 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	Rumania	Law
4	1938	Cerban	Alexandru	University of Bukarest	Rumania	Law

The research assistants would consider lines 1 and 2 as linked. In contrast, they would not classify lines 3 and 4 as linked. To decide whether two lines are linked, the research assistants only allow for very minor differences in the spelling of the first name, such as Harrison Mc Allister and Harrison McAllister.

Across departments in the same country

In a second step, we link academics who remain in the same country but change departments between two Minerva cohorts. We link all academics from discipline d, country c, cohort m to all academics from the same discipline d, same country c but cohort $m+1$ based on the academic's surname, the first initial, her discipline (e.g., physics, geography, and so on) and her country. Hence, all potential links that we consider have identical surnames, 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.[44] We then link academics across Minerva cohorts as follows:

[44]A small number of universities change countries 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 merge for the University of Strasbourg considers academics who move from or to other German universities between 1900 and 1914. It also considers academics who move from or to other French universities between 1925 and 1938, 1938 and 1952, and 1952 and 1966. The moves between 1914 and 1925 (when the university changes country) are considered in the cross-country merge we describe below.

1. If the entire information on the first name (in some cases the information on the first name that Minerva reports for that particular academic may be one or more initials in both Minerva cohorts) is identical in both Minerva cohorts, we consider these academics as linked across the two Minerva cohorts.

2. If the information on the first name differs across the two Minerva cohorts, research assistants examine each potential link and decide whether the academics can actually be linked. To decide whether a potential link is valid, the research assistants use the following rules:

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

   (b) If all initials of the first name are identical and if the first name contains more than one initial (even if the first name differs) the potential link is classified as an actual link (see lines 3 and 4 in Table A.2)

   (c) If only one initial is reported for one Minerva cohort but a full first name in the other Minerva cohort the research assistants google the relevant scientist. If the research assistants find online biographical information that confirms that the academic was indeed employed at university u in the year corresponding to Minerva cohort t and then moved to university μ before the year corresponding to Minerva cohort $t+1$, the potential link is classified as valid

   E.g., K(arl) Röder (see lines 5 and 6 in Table A.2) could be found online (see Wikipedia Article) and his Wikipedia entry states that:

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

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

Table A.2: Examples: Within Country Merge

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

Across countries In a third step, we link academics who move across countries. We link all academics from discipline d and cohort t to all academics from the same discipline d but cohort $t+1$ who are listed in two different countries. As such links may be more likely to be false positives, all potential links are confirmed by extensive manual online searches.

Potential links are based on the academic's surname, the first initial, her discipline (e.g., physics, geography, and so on). Hence, all potential links that we consider have identical surnames and disciplines, but are listed in different countries (and hence different universities) 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 Minerva cohort t and then moved to university μ in country c' before the year corresponding to Minerva cohort $t+1$, the potential link is classified as valid.

8.1.5 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.[45] The raw data report full first names for about 77% of academics. For the remaining 23% of academics, Minerva only lists initials. We increase the share of academics with full first names in two ways. First, we use information on the same academic from a different cohort (see 8.1.4). E.g., the University of Chicago botanist *Henry Chandler* Cowles was reported as *Henry C.* Cowles in 1914 but as H. C. Cowles in 1925. We therefore adjust the first name in 1925 to Henry C. Second, we hand-check around 60,000 academics who are only reported with initials. For this step, research assistants google the initial(s), surname, discipline, and university to find online records for the respective academic.

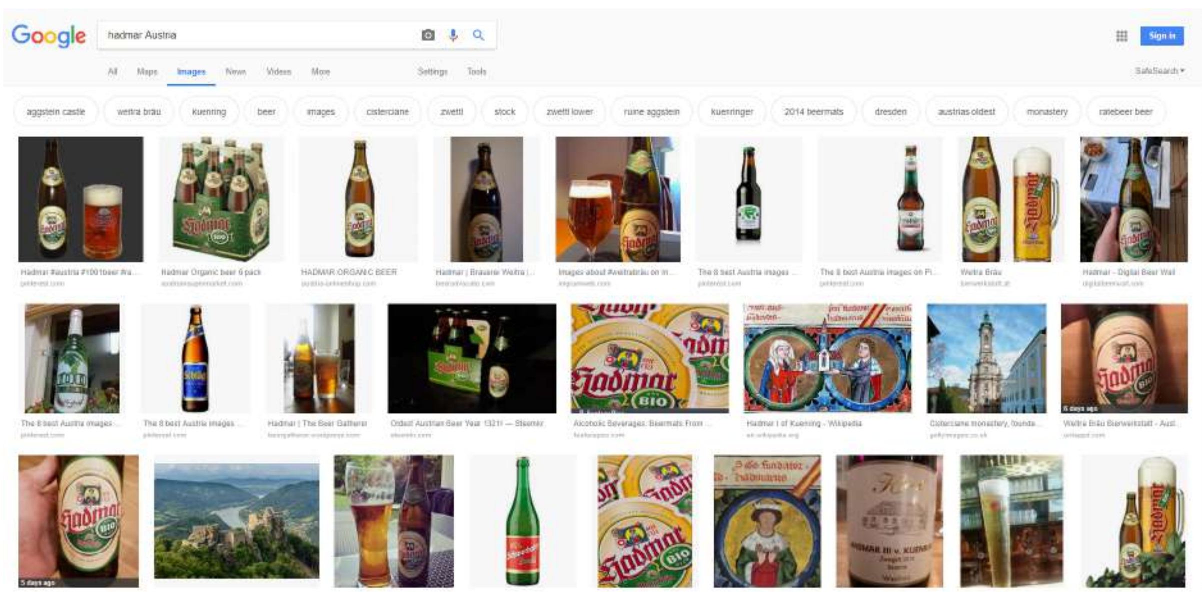
[45] For a small number of academics we can use information on gender from the way that academics are listed in Minerva (e.g., as Miss or Mlle.)

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.

8.1.6 Additional Information on Coding Gender

Example Google Picture Search As described in the main text, one of the steps to identify the gender of academics relies on a google picture search for first name by country combinations. Figure A.1 shows an example of the output of the google picture search, when we searched for "Hadmar Austria."

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



Notes: The Figure shows an example of the Google picture 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 picture search is used if gender-api.com and the hand-coding of research assistants cannot assign gender to a first name by country combination (see section 2.1 for details).

Hand-Checking Gender Coding As described in the main text, in the last step of the gender assignment, we hand-check individual academics who appear mis-classified. Such mis-classifications occur mostly because the predominant gender of a first name by country combination changes over time.[46] 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

[46]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.

potential of mis-classification (e.g., Camille in France). Second, research assistants google the actual scientist and try to establish their gender. E.g., for the French biologist Camille Sauvageau, they find an entry in the *Proceedings of the Linnean Society of London* (from 1937) which says: "Camille Savageau (1861-1936), Foreign Member of the Society, was born in Angers on 12 May 1861. *He* studied at Montpellier..." (see Link Google Books for details).

8.2 Preparation of Web of Science Data

8.2.1 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."[47] 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 *Minerva*, we go through the following steps to homogenize the full scientist name as reported by the *Web of Science* to the corresponding information in *Minerva*:

1. We remove titles such as "Jr." or "Dr." or "Prof." from the full scientist 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 nobel titles, e.g., "Della" or "Op Den" or "Von Der" or "Viscount."

4. We extract initial(s) from the scientist's first name(s).

5. We further extract the first of the initials from the list of first name(s) initial(s) obtained at the previous step.

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

8.2.2 Addresses of Papers

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 actual lists an address. In some of these cases, an alternative database, *Microsoft Academic Graph* (*MAG*), contains the relevant address information. We therefore enrich the scientists' addresses as reported by the *Web of Science* with information from *MAG*.[48] 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 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 address 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 observe an address. We then assign the address, St. Louis, USA, from the first paper to the second paper.

Processing Addresses with Google Maps

Over the very long time period that we study in this paper, some cities change their name (e.g. St. Petersburg became Leningrad) and a number of cities change countries (e.g. Strasbourg was German in 1900 and 1914, and then became French for the later cohorts). Furthermore,

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

cities may be spelled in different languages in Minerva and on a paper in the *Web of Science*. E.g. Rome is spelled using the German spelling "Rom" in *Minerva* but spelled with either Italian ("Roma"), English ("Rome"), or German ("Rom") spelling in the Web of Science, depending on the country of the journal. To allow the match of papers to the academics in *Minerva* we therefore harmonize the address (in particular the country and city) in the *Web of Science* with the address in *Minerva* 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*.[49] *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 which indicates the type of address that has been found (e.g. "CITY" if the *Google API* found a city). Similarly, we geocode the *Minerva* data with *Google Maps API*. This also returns updated names of cities and countries. Crucially, as we process both addresses in the *Web of Science* and in *Minerva* with *Google Maps API* we obtain a harmonized set of addresses without spelling inconsistencies.

Step 1, part ii) In some cases, Google does not find the correct city and country. This usually occurs either because the names of a city or a country has changed over time (e.g. the name of Preßburg changed into Bratislava) or because of typos in the *Web of Science*. These cases are easily identifiable as the location type flag is "APPROXIMATE" instead of "CITY". We improve the geocoding for these cases using the following 3-step procedure:

1. We structure the address before submitting it to the API, (e.g. "'city' : Preßburg, 'country' : Hungary").[50]

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

3. For those cases that did not return a result in steps one and two, we submit the complete address from the *Web of Science* (e.g. "Loyola Univ Clinics, Mercy Hosp, Chicago, IL USA") to the API.

Step 2 In some cases, the procedure above does not guarantee that the correct city and country has been found. We therefore rely on the *Google Maps web interface*, as opposed

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

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

to the *API,* for the second step to improve the address data for addresses that appear mis-classified. 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 adresses that are mis-classified, we calculate the Levenshtein distance between the city name in the *Web of Science* and the city name that Google returned. If the Levenshtein distance is larger than three (i.e., more than 3 letters differ), we copy the full address from the *Web of Science* into the *Google Maps* web interface. If the web interface finds the address, we extract the city and country information from the website and use them as inputs for the *Google Maps API* (i.e. *Step 1, part i*). We further process the output from the *Google Maps web interface* with Google Maps API because the web interface returns somewhat different city and country names than the API.

The processing of addresses ensures that addresses in *Minerva* and the *Web of Science* are harmonized and can then be matched as described in subsection 8.3 below.

8.2.3 Predicting Academic Disciplines of Papers Using Paper Titles

To match papers from the *Web of Science* to *Minerva* we also match on the discipline (subsection 8.3 below). The *Web of Science* assigns papers to academic disciplines (e.g. physics, or general science) on the basis of the journal they are published in, as opposed to assigning each individual paper to a unique discipline. For 59% of the papers, this establishes a unique assignment to one of 8 disciplines. The remaining 41% papers are published in journals that the *Web of Science* either assigns to multiple disciplines (e.g., the journal *Biometrika* is assigned to mathematics as well as biology) or to general science (e.g., *Nature* and *Science*).[51] Matching these papers to academics in Minerva would involve considerable measurement error.

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

In preparation for the classifier, we remove very common words (stopwords) from the titles, as these do contain very 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

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

the size of the vocabulary V is the total number of unique unigrams, bigrams, and trigrams in all titles.

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

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

The multinomial logistic regression classifier then learns to predict disciplines based on the 1,2,3-gram matrix $\mathbf{X}$, where the dependent variable y_d is the discipline of the paper. To avoid overfitting, we include L2 regularization in the classifier. As is standard in the machine learning literature, the optimal regularization strength is chosen using 10-fold cross-validation and evaluated on the basis of the F1-score.[52] The final classifier achieves a within-sample F1-score of 0.99 and an out-of-sample F1-score of 0.81. After the training process, we predict a unique discipline for the 10,508,299 papers which the Web of Science had originally assigned to multiple disciplines (on the basis of the journal).

8.3 Merging Minerva to Web of Science

We match papers from the *Web of Science* to the data on scientists from Minerva using a six-step matching procedure. As mentioned in the main text, we only match papers from the *Web of Science* within a ± five-year-window around the year of the corresponding *Minerva* cohort. E.g., for scientists listed in *Minerva* 1914, we only match papers published between 1909 and 1919.[53] Within these windows, we match the *Web of Science* data to each cohort of the *Minerva* data on the basis of 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

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

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

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 relevant scientist listed in *Minerva*. In particular, we use the following rule to exclude false matches: Denote the string of initials of a scientist in *Minerva* by s and that 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 scientist listed in Minerva with initials $A.A.$ will not be merged to a paper published by someone with with initials $A.B.$ (Note: as described under steps 1-3, we only consider matches where the full surname, subject, country, and city matches.)

   (b) If the number of initials in s and p is not identical ($|s| \neq |p|$), we exclude matches in which not all letters from the shorter set of initials appear in the other. 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 gets excluded. For example, a scientist listed in *Minerva* with initials $A.B.$ will not be merged to a paper published by someone with $A.C.D.$ or $A.C.B.$

4. We then repeat steps 1-4, but removing the city as a merge criterion.

5. We repeat steps 1-4, but additionally removing country as a merge criterion.

Note: if one of the authors of a paper is matched to a scientist in Minerva in an earlier (and thus more restrictive) step, this particular author will no longer be considered in any further step. We account for the fact that some papers are merged to multiple scientists from the *Minerva* data by weighting the papers by the total number of matches. For the time period covered by our paper, the *Web of Science* rarely provides a unique assignment of the addresses reported on a paper to its co-authors: e.g., if a paper has two co-authors and these are affiliated to different institutions, usually the *Web of Science* does not pin down which co-author is affiliated to which institution. We therefore merge each address reported in a paper to all of the co-authors of the paper. If there is more than one address associated to a paper, we perform a many to many merge of addresses to co-authors.

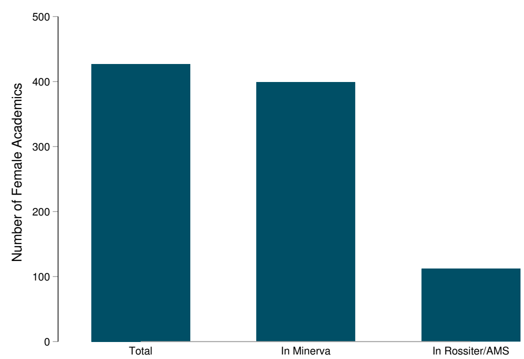
8.4 Benchmarking the Minerva Data

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

8.4.1 Benchmarking Against Rossiter (1982) / American Men of Science (1938)

Rossiter (1982), pp. 182 reports female scientists in twenty major U.S. universities for the year 1938. The data are based on women listed in the historical publication *American Men of Science (AMS),* 6th edition, 1938. The data contain all female scientists who are listed in the *AMS* for twenty leading U.S. institutions. For the benchmarking exercise, we extract all female scientists who are at least assistant professors that are listed in these twenty universities in *Minerva* 1938. We then cross-check all names and identify women listed in both sources. Both sources combined list a total of 427 different female academics which we take as the best available information for the total number of women in these twenty universities in 1938 (first bar, Figure A.2). Of these, 399 (93%) are listed in *Minerva* (second bar).[54] In contrast, Rossiter on the basis of the *American Men of Science* only lists 112 (26%) of them (third bar). This indicates that Minerva 1938 has a much more comprehensive coverage of academics in the top twenty U.S. universities for 1938 than the *American Men of Science.*

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



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

[54]The 7% missing female academics in Minerva most likely come from the following reasons: 1) in 1938 Minnesota (one of the 20 universities) only reported full professors but Rossiter reports 9 female assistant or associate professors in Minnesota. 2) even though both sources were published in 1938 they may report faculty from slightly different cutoff dates.

8.4.2 Benchmarking Against German University Catalogue 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*. We extract all physicists, chemists, and mathematicians in the same way as Waldinger 2012a.

Overall, these data contain 866 scientists in the three fields for the winter semester 1937/38. We then match these scientists to *Minerva*, matching on the surname, first name, discipline, and university. Of the 866 scientists we are able to match 853 scientists, a match rate of 98.5 percent, suggesting that the coverage of *Minerva* was very comprehensive (see Figure A.3).

Figure A.3: Benchmarking Minerva Against Faculty Rosters for German University 1938

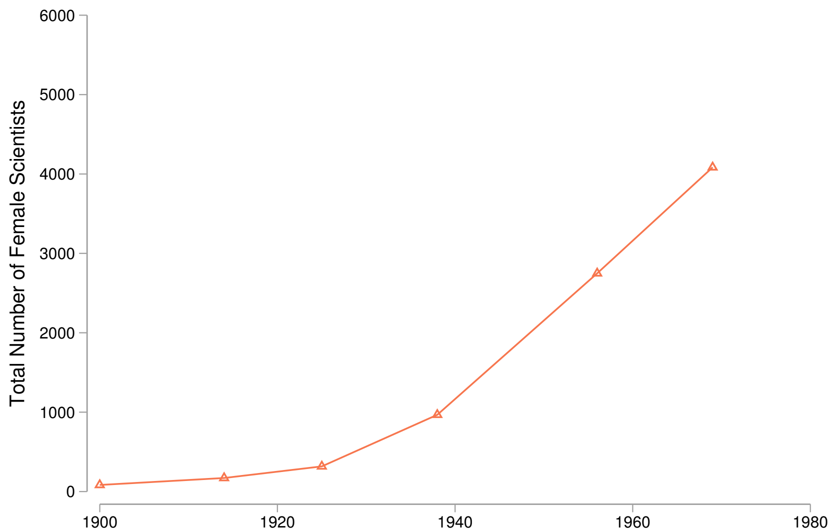
Notes: The Figure shows the total number of scientists in the *Kalender der reichsdeutschen Universitäten und Hochschulen 1937/38* and how many of these could be found in *Minerva 1938.*

9 Further Results: Hiring Gaps

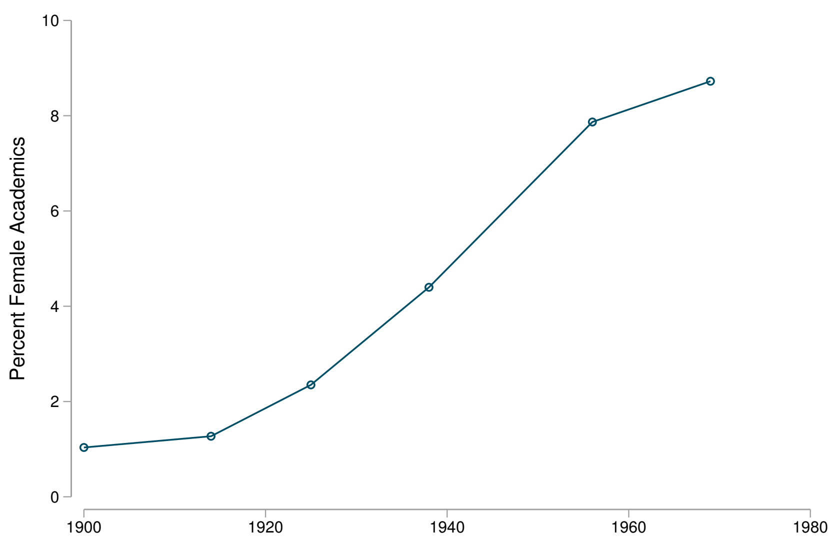
The following figures provide additional results on hiring gaps.

Figure B.1: Female Academics Over Time: Science Disciplines

(a) Total Number of Female Scientists



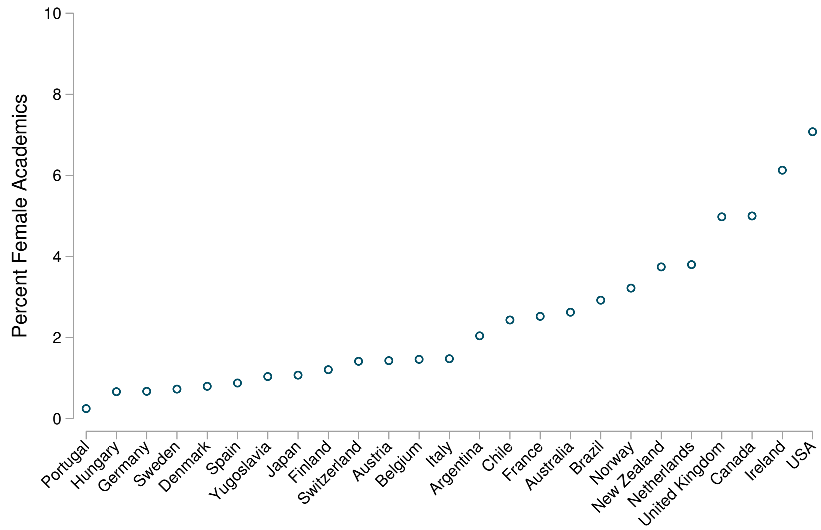
(b) Percent of Female Scientists



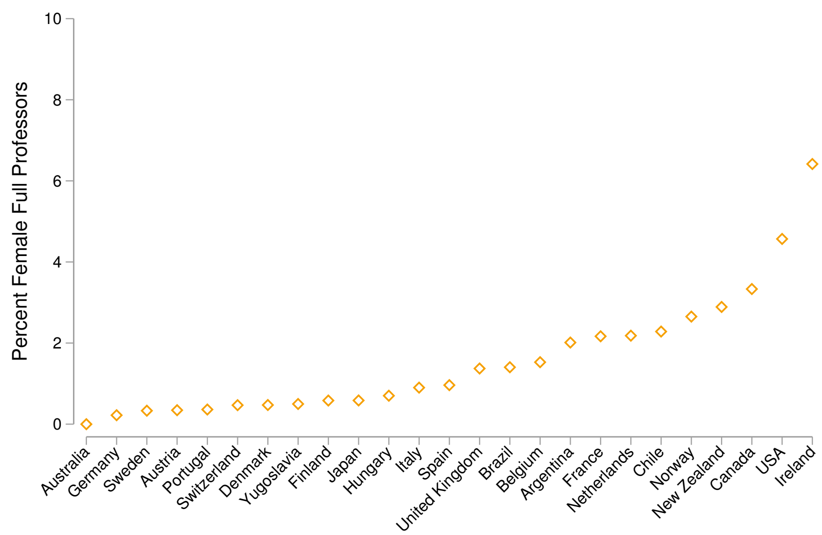
Notes: The Figure shows the total number and share of female scientists in the six Minerva cohorts. Panel (a) shows the total number of female scientists of all ranks, where scientists are defined as academics working in mathematics, physics, chemistry, biochemistry, biology, and medicine. Panel (b) shows the share of female scientists of all ranks. The data were collected by the authors from various volumes of Minerva, see section 2 for details.

Figure B.2: Percent of Female Academics by Country, Stable Set of Universities

(a) All Ranks



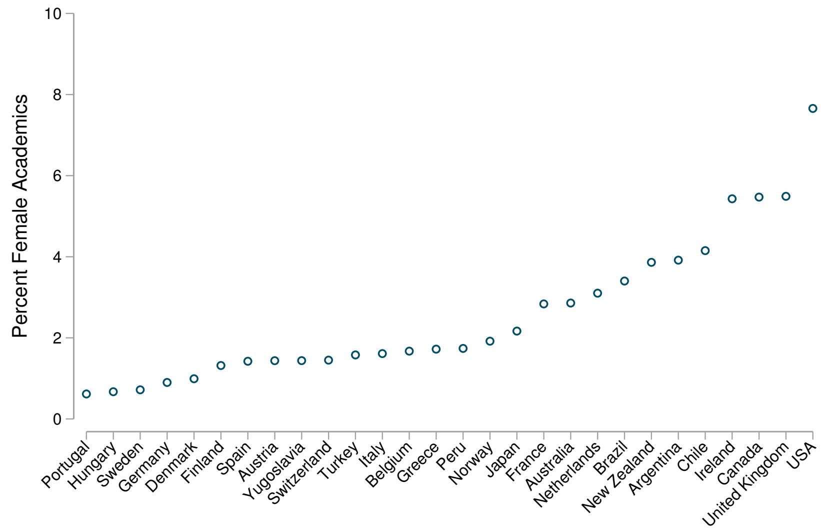
(b) Percent of Female Full Professors



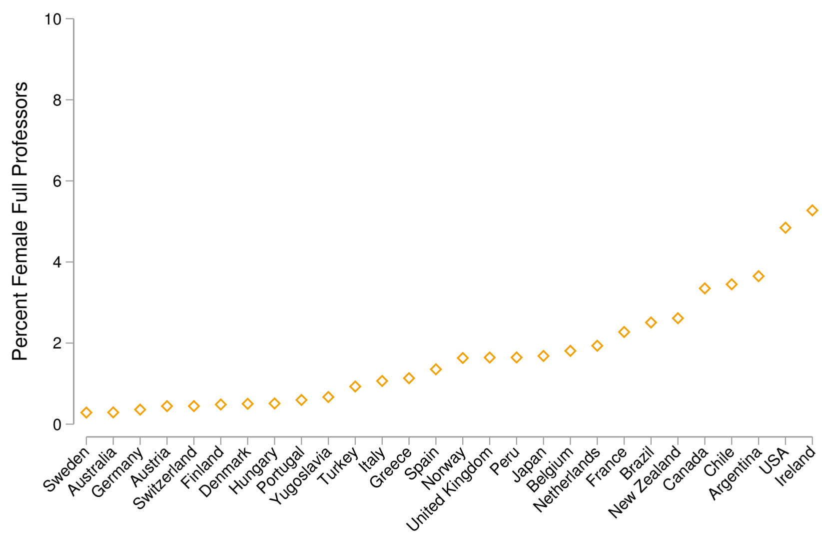
Notes: The Figure shows the percentage of female academics by country, as computed from universities observed in all six cohorts. Panel (a) shows academics of all ranks. Panel (b) shows full professors, only. We calculate percentages of female academics at the cohort and country level, e.g. United States in 1900 or United States in 1914, and then average the percentages over the six cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort). The data were collected by the authors from various volumes of Minerva, see section 2 for details.

Figure B.3: Percent of Female Academics by Country, Excluding Women's Colleges

(a) All Ranks



(b) Percent of Female Full Professors



Notes: The Figure shows the percentage of female academics by country, excluding from the sample the women's colleges. Panel (a) shows academics of all ranks. Panel (b) shows full professors, only. We calculate percentages of female academics at the cohort and country level, e.g. United States in 1900 or United States in 1914, and then average the percentages over the six cohorts (so that each cohort receives the same weight, independently of the total number of academics in that cohort). The data were collected by the authors from various volumes of Minerva, see section 2 for details.

10 Further Results: Publication Gaps

10.1 Additional Results: Publication Gaps

Table C.1: Individual-Level Publication Gaps (Unique Matches)

Dependent Variable	(1) Publi-cations	(2) Publi-cations	(3) Publi-cations	(4) Publi-cations	(5) Publi-cations	(6) Standard. Publi-cations	(7) Standard. Publi-cations	(8) Standard. Publi-cations	(9) Standard. Publi-cations	(10) Standard. Publi-cations
Panel A: Full Set of Universities										
Female	-2.572***	-2.703***	-2.186***	-2.250***	-2.189***	-0.233***	-0.236***	-0.177***	-0.186***	-0.181***
	(0.556)	(0.606)	(0.626)	(0.639)	(0.655)	(0.020)	(0.018)	(0.021)	(0.024)	(0.028)
Observations	116,441	116,441	116,441	116,441	116,441	116,441	116,441	116,441	116,441	116,441
R-squared	0.088	0.134	0.184	0.220	0.264	0.004	0.006	0.069	0.101	0.157
Panel B: Stable Set of Universities										
Female	-3.345***	-3.599***	-3.308***	-3.320***	-3.253***	-0.327***	-0.320***	-0.282***	-0.278***	-0.268***
	(0.564)	(0.599)	(0.613)	(0.574)	(0.638)	(0.029)	(0.029)	(0.019)	(0.017)	(0.025)
Observations	51,503	51,503	51,503	51,503	51,503	51,503	51,503	51,503	51,503	51,503
R-squared	0.107	0.151	0.179	0.200	0.245	0.011	0.019	0.051	0.068	0.124
Cohort FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country FE	Yes					Yes				
Cohort × Discipline × Country FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist-level, where each scientist is defined as a unique lastname - first initial - discipline combination in every cohort. Panel A reports results for scientists from all universities, while panel B reports results for scientists from universities observed in all six cohorts. In columns 1-5, the dependent variable equals the number of publications in a ±5 - year window around a Minerva cohort (i.e. 1909-1919 for a scientist listed in 1914). In columns 6-10, the dependent variable equals the number of publications, but standarized at the country-cohort-discipline level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman. The regressions also control for various fixed effects, as indicated in the table. Standard errors are clustered at the discipline-country level. Significance levels: ∗∗∗ p<0.01, ∗∗ p<0.05, and ∗ p<0.1.

Table C.2: Individual-Level Publication Gaps (±Three-Year Window)

Dependent Variable	(1) Publi-cations	(2) Publi-cations	(3) Publi-cations	(4) Publi-cations	(5) Publi-cations	(6) Standard. Publi-cations	(7) Standard. Publi-cations	(8) Standard. Publi-cations	(9) Standard. Publi-cations	(10) Standard. Publi-cations
Panel A: Full Set of Universities										
Female	-1.477***	-1.620***	-1.310***	-1.349***	-1.316***	-0.204***	-0.209***	-0.159***	-0.167***	-0.165***
	(0.255)	(0.290)	(0.300)	(0.285)	(0.286)	(0.013)	(0.013)	(0.013)	(0.014)	(0.017)
Observations	132,677	132,677	132,677	132,677	132,677	132,677	132,677	132,677	132,677	132,677
R-squared	0.080	0.118	0.158	0.189	0.223	0.002	0.002	0.055	0.085	0.133
Panel B: Stable Set of Universities										
Female	-1.867***	-2.049***	-1.882***	-1.890***	-1.816***	-0.279***	-0.272***	-0.245***	-0.241***	-0.232***
	(0.209)	(0.214)	(0.224)	(0.202)	(0.207)	(0.026)	(0.025)	(0.015)	(0.015)	(0.012)
Observations	69,147	69,147	69,147	69,147	69,147	69,147	69,147	69,147	69,147	69,147
R-squared	0.096	0.131	0.154	0.175	0.213	0.007	0.012	0.041	0.059	0.110
Cohort FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country FE	Yes					Yes				
Cohort × Discipline × Country FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes

Notes: The Table shows gender gaps in publications. Results are estimated at the scientist-level. Panel A reports results for scientists from all universities, while panel B reports results for scientists from universities observed in all six cohorts. In columns 1-5, the dependent variable equals the number of publications in a ±3 - year window around a Minerva cohort (i.e. 1911-1917 for a scientist listed in 1914). In columns 6-10, the dependent variable equals the number of publications, but standarized at the country-cohort-discipline level. The main explanatory variable is an indicator that equals 1 if the scientist is a woman. The regressions also control for various fixed effects, as indicated in the table. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

10.2 A Stylized Model Of Gender Gaps

We develop a stylized model connecting the publication gaps and the share of female academics at the country-level, rationalizing the "gender swoosh" in Figure 10a. The proposed model builds on Roy (1951) and allows for (i) selection on unobservables in the hiring market, (ii) gender bias in hiring, and (iii) gender bias in the publication market. These features lead a scientist's publication outcome to be a function of the share of women in a country-period because of (a) *indirect* effects of selection and gender bias in the hiring market and (b) the *direct* effects of gender bias in the publication market.

First, at the hiring stage, denoted by 0, for any available academic position i, a woman W or a man M can be hired: in case of any gender imbalance, we refer to the gender bias in hiring as Δ_0. Second, for all hired academics, we observe publication outcomes at a later stage, denoted by 1, and we refer to any gender bias in publications as Δ_1. The model allows for the possibility that selection at the hiring stage 0 may affects the observed publications at stage 1, so that e.g. the gender gap in publications is indirectly a function of Δ_0 and not only of Δ_1.

The model accounts for the fact that we cannot observe the relevant population of individuals who could have potentially found a job in academia. Even if world-wide data on

PhD graduates from all disciplines were available, these would suffer from a further form of endogeneous selection as women were initially barred from entering PhD programs basically everywhere (Rossiter 1982). Our data however allow us to observe the share of women in academia $s_0^W = \frac{\Pr[W]}{\Pr[W]+\Pr[M]}$, where $\Pr[W]$ denotes the probability that a woman enters academia and $\Pr[M]$ denotes the probability that a man enters academia, and the Roy model we propose builds on observing the share of women who have been hired, s_0^W.

To provide intuition, we start by introducing a simplified version of the model that relies on the following assumptions:

(i) Δ_0 is not a function of s_0^W,

(ii) Δ_1 is not a function of s_0^W.

In subsection 10.2.2, we then present and estimate a more general version of the model that relaxes assumption (ii) and allows the gender gap in publications to be affected by the share of women in academia.

10.2.1 A Simplified Model

Selection in Hiring Market. Suppose that there is an academic position i that will be allocated to either a woman W or a man M. The expected value of hiring a woman is:

$$Y_{0i}^W = X_i^W \beta_0 + \epsilon_{0i}^W, \tag{C.1}$$

while that of hiring a man is:

$$Y_{0i}^M = X_i^M \beta_0 + \Delta_0 + \epsilon_{0i}^M, \tag{C.2}$$

where X_i^g, $g \in \{W, M\}$, are observable characteristics, Δ_0 a possible gender bias in hiring, and ϵ_{0i}^g is the unobserved component of these expectations. As a result, academic position i is given to a woman if (C.1) is greater than (C.2):

$$\begin{aligned} Y_{0i} &= \left(X_i^W - X_i^M\right)\beta_0 - \Delta_0 + \left(\epsilon_{0i}^W - \epsilon_{0i}^M\right) \\ &= X_i\beta_0 - \Delta_0 + \epsilon_{0i} \qquad \text{(C.3)} \\ &> 0, \end{aligned}$$

so that, when $\Delta_0 > 0$ but everything else is equal, women need to overcome the additional hurdle or gender bias Δ_0 to be hired in position i over men. Assuming that ϵ_{0i} is distributed i.i.d. normal (assumption (iv) below), the probability that a woman is hired in academic position i is:

$$\begin{aligned} s_{0i}^W = \quad & \Pr[Y_{0i} > 0] = \Pr[\epsilon_{0i} > -X_i\beta_0 + \Delta_0] \\ & = \Phi\left(X_i\beta_0 - \Delta_0\right), \end{aligned} \tag{C.4}$$

from which it follows that:

$$\Phi^{-1}\left(s_{0i}^W\right) = X_i\beta_0 - \Delta_0. \tag{C.5}$$

The differences in observable characteristics X_i could be based for example on SAT scores or college GPA. Such data are unfortunately not available on a world-wide scale covering the twentieth century. Then, in our application $X_i = 0$ and equation (C.5) simply reduces to $\Phi^{-1}\left(s_0^W\right) = -\Delta_0$, where s_0^W is the share of women among all academics. In this case, as we observe the share s_0^W in the data, we can directly compute Δ_0 without performing any estimation.

Publication Market. Conditional on academic position i being allocated to a woman or a man at the hiring stage, we have the following outcome equation at the publication stage:

$$\begin{aligned} Y_{1i}^W &= Z_i^W\beta_1 + \epsilon_{1i}^W & \text{if } Y_{0i} > 0 \\ Y_{1i}^M &= Z_i^M\beta_1 + \Delta_1 + \epsilon_{1i}^M & \text{if } Y_{0i} \leq 0, \end{aligned} \tag{C.6}$$

where Z_i^g, $g \in \{W, M\}$, are observable characteristics and ϵ_{1i}^g is the unobserved component of the publication outcome Y_{1i}^g. In words, if academic position i is given to a woman, we observe the publication outcome of a woman, otherwise we observe the publication outcome of a man. Since for any i we cannot observe the counterfactual publication outcome (i.e., the number of publications if position i had been allocated to the other gender), (C.6) will suffer from selection on unobservables if the error terms in (C.3) and (C.6) are correlated: e.g., talented women, such as Marie Curie, were both more likely to get academic positions and to publish well once hired.

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

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

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

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

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

$$\begin{aligned} \mathbb{E}\left[Y_{1i}^{W}\middle| X_i, Z_i^W, Y_{0i} > 0\right] &= Z_i^W \beta_1 + \mathbb{E}\left[\rho_W \epsilon_{0i} + \xi_i^W \middle| \epsilon_{0i} > -X_i\beta_0 + \Delta_0\right] \\ &= Z_i^W \beta_1 + \rho_W \lambda\left(X_i\beta_0 - \Delta_0\right), \end{aligned} \tag{C.7}$$

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

$$\mathbb{E}\left[Y_{1i}^{M}\middle| X_i, Z_i^M, Y_{0i} \leq 0\right] = Z_i^M \beta_1 + \Delta_1 - \rho_M \lambda\left(\Delta_0 - X_i\beta_0\right). \tag{C.8}$$

Putting together (C.7) and (C.8) closes the model. When the gender bias in publications Δ_1 increases (e.g., because of biased editors or referees Card et al. 2020a), the conditional expectation of publications Y_{1i}^{M} increases. When the gender bias in hiring Δ_0 increases, i.e. only the best women are hired, the inverse Mills ratio $\lambda(\cdot)$ goes up, and the conditional expectation of publications Y_{1i}^{W} increases if $\rho_W > 0$. This holds independently of any gender bias Δ_1 in publications. In contrast, equation (C.8) indicates that when the gender bias in hiring Δ_0 increases, because $\lambda(\cdot)$ approaches zero, men's conditional expectation of Y_{1i}^{M} may be unaffected, even if $\rho_M \neq 0$. This highlights how selection and gender biases in hiring can indirectly affect estimated gender gaps in publications (in addition to any direct gender bias in publications). An advantage of this model is that the inverse Mills ratio does not depend on Z_i^W and Z_i^M but only on the observed share of women (an information otherwise not used in the publication outcome equation), which serves the purpose of an exclusion restriction.

10.2.2 A More General Model

In the more general version of the model that we estimate, we relax assumptions (ii)-(iv). First, we allow the gender bias in publications Δ_1 to be a function of the share of female academics, $\Delta_1\left(s_0^W\right)$. Furthermore, we generalize hiring equation (C.3) to the semiparametric specification:

$$Y_{0i} = \epsilon_{0i} - r_0\left(X_i\right), \tag{C.9}$$

with $r_0(\cdot)$ an unknown nonparametric function which incorporates Δ_0[55] and the error term ϵ_{0i} distributed according to F, an unknown CDF with unbounded support and invertible. Because in our application $X_i = 0$, so that $r_0(X_i) = \Delta_0$, the share of women academics can simply be expressed as:

$$\begin{aligned} s_0^W &= \Pr[Y_{0i} > 0] = \Pr\left[\epsilon_{0i} > \Delta_0\right] = 1 - \Pr\left[\epsilon_{0i} \leq \Delta_0\right] \\ &= 1 - F\left(\Delta_0\right), \end{aligned} \tag{C.10}$$

while the share of men academics is $1 - s_0^W = F(\Delta_0)$. Given this model and assumptions, the expectation of Y_{1i}^W conditional on $Y_{0i} > 0$ from (C.6) is:

$$\begin{aligned} \mathbb{E}\left[Y_{1i}^W \middle| X_i, Z_i^W, Y_{0i} > 0\right] &= Z_i^W\beta_1 + \mathbb{E}\left[\epsilon_{1i}^W \middle| \epsilon_{0i} > \Delta_0\right] \\ &= Z_i^W\beta_1 + \widetilde{g}_W\left(\Delta_0\right) = Z_i^W\beta_1 + \widetilde{g}_W\left(F^{-1}\left(1 - s_0^W\right)\right) \\ &= Z_i^W\beta_1 + g_W\left(s_0^W\right), \end{aligned} \tag{C.11}$$

where $g_W\left(s_0^W\right) = \widetilde{g}_W \circ F^{-1}\left(1 - s_0^W\right)$ is the nonparametric counterpart of $\rho_W \lambda \circ \Phi^{-1}\left(s_0^W\right)$ in (C.7) (when $X_i = 0$). Analogously, the expectation of Y_{1i}^M conditional on $Y_{0i} \leq 0$ from (C.6) is:

[55] We assume that, while nonparametric with respect to X_i, $r_0(\cdot)$ must however satisfy the exclusion restriction embodied in assumption (i): $r_0(\cdot)$ cannot depend on s_0^W. For example, $r_0(\cdot)$ trivially fits the linear specification from equation (C.3), $r_0(X_i) = X_i\beta_0 + \Delta_0$, but can also take the more general form $r_0(X_i) = g(X_i, \beta_0) + \Delta_0$, with $g(X_i, \beta_0)$ any function of X_i and the finite- or even infinite-dimensional parameter β_0.

$$
\begin{aligned}
\mathbb{E}\left[Y_{1i}^M \middle| X_i, Z_i^M, Y_{0i} \leq 0\right] &= Z_i^M \beta_1 + \Delta_1\left(s_0^W\right) + \mathbb{E}\left[\epsilon_{1i}^M \middle| \epsilon_{0i} \leq \Delta_0\right] \\
&= Z_i^M \beta_1 + \Delta_1\left(s_0^W\right) + \tilde{g}_M\left(F^{-1}\left(1 - s_0^W\right)\right) = Z_i^M \beta_1 + \Delta_1\left(s_0^W\right) + g_M\left(s_0^W\right) \\
&= Z_i^M \beta_1 + G_M\left(s_0^W\right),
\end{aligned} \tag{C.12}
$$

where again $G_M\left(s_0^W\right) = \Delta_1\left(s_0^W\right) + \tilde{g}_M \circ F^{-1}\left(1 - s_0^W\right)$ is the nonparametric counterpart of $\Delta_1 - \rho_M \lambda \circ \left[-\Phi^{-1}\left(s_0^W\right)\right]$ in (C.7) (when $X_i = 0$).

Without further assumptions, it is not possible to separately identify the various components of $g_W(\cdot)$ in equation (C.11) and $G_M(\cdot)$ in equation (C.12).[56] To avoid unnecessarily strong functional form restrictions, we directly approximate these functions by polynomial expansions of the share of female academics:[57]

$$
\begin{aligned}
g_W\left(s_0^W\right) &= \sum_{\kappa=0}^{K} \theta_\kappa^W \times \left(s_0^W\right)^\kappa \\
G_M\left(s_0^W\right) &= \sum_{\kappa=0}^{K} \theta_\kappa^M \times \left(s_0^W\right)^\kappa,
\end{aligned} \tag{C.13}
$$

where K is the degree of the polynomial. For finite K, the resulting estimator of (C.11) and (C.12) would be parametric (an OLS), while for $K \to \infty$ as the number of observations grows, the estimator would be a nonparametric sieve. Our data enable us to compute the hiring share $s_{0\ell}^W$ at the country and time period level, where $\ell = 1, ..., L$ denotes countries as well as time periods. Suppose for instance that $K = 3$. Then, combining (C.6), (C.11), and (C.12), we obtain the publication outcome equations:

$$
\begin{aligned}
Y_{1i}^W &= Z_i^W \beta_1 + \sum_{\kappa=0}^{3} \theta_\kappa^W \times \left(s_{0\ell_i}^W\right)^\kappa + e_{1i}^W \quad \text{if } Y_{0i} > 0 \\
Y_{1i}^M &= Z_i^M \beta_1 + \sum_{\kappa=0}^{3} \theta_\kappa^M \times \left(s_{0\ell_i}^W\right)^\kappa + e_{1i}^M \quad \text{if } Y_{0i} \leq 0,
\end{aligned} \tag{C.14}
$$

[56]Assumptions (ii)-(iv) would imply the separate identification of these components uniquely via functional form restrictions.

[57]More broadly, any series expansion of these functions could be used, e.g. spline sieves.

where $e_{1i}^W = \epsilon_{1i}^W - g_W\left(s_{0\ell_i}^W\right)$ and $e_{1i}^M = \epsilon_{1i}^M - G_M\left(s_{0\ell_i}^W\right)$ have both zero conditional expectation if equation (C.13) holds (i.e., we control for any endogenous sample selection term). We can estimate the parameters in (C.14) from the following OLS regression:

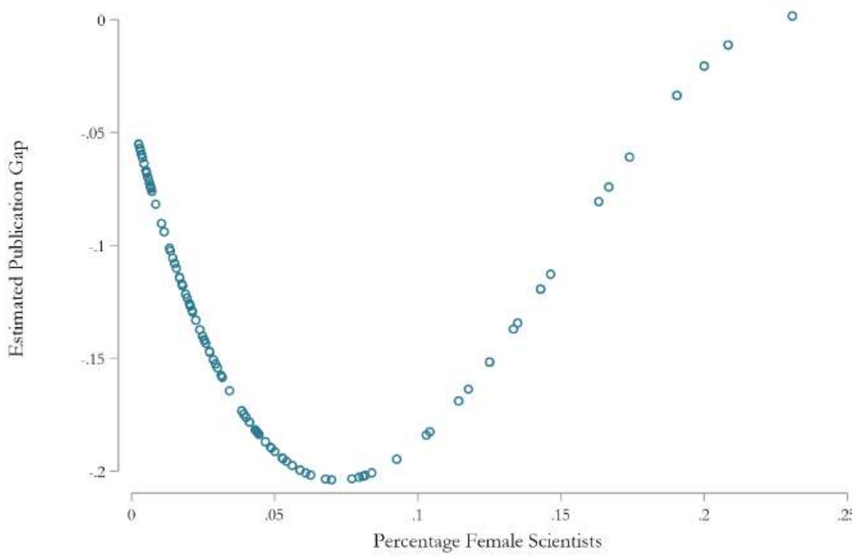
$$\begin{aligned} Y_{1i}^M + \text{Female}_i\left(Y_{1i}^W - Y_{1i}^M\right) &= \left[Z_i^M + \text{Female}_i\left(Z_i^W - Z_i^M\right)\right]\beta_1 + \sum_{\kappa=0}^{3}\theta_\kappa^M \times \left(s_{0\ell_i}^W\right)^\kappa \\ &+ \sum_{\kappa=0}^{3}\left(\theta_\kappa^W - \theta_\kappa^M\right)\text{Female}_i \times \left(s_{0\ell_i}^W\right)^\kappa + FE_i + \varepsilon_i, \end{aligned} \tag{C.15}$$

where Female_i is a female dummy that denotes whether $Y_{0i} > 0$ for each observation i, FE_i are various fixed effects, and ε_i is a residual term. In practice, we estimate (C.15) on the basis of the scientist sample using a similar regression to (1):

$$\text{Pub}_{idt} = \gamma + \sum_{\kappa=0}^{3}\gamma_\kappa\,\text{Female}_{idt} \times \left(s_{0\ell_i}^W\right)^\kappa + \text{Fixed Effects} + \varepsilon_{idt}, \tag{C.16}$$

where Pub_{idt} measures the standardized number of papers published by scientist i in cohort t and department d, the fixed effects are at the level of the cohort $\times$ department, and each γ_κ corresponds to $\theta_\kappa^W - \theta_\kappa^M$, $\kappa = 0, ..., 3$. Similar to the results presented in Figure 10a, we compute $s_{0\ell_i}^W$ at the country-pre/post WWII level and, hence, in regression (C.16) the terms $\sum_{\kappa=0}^{3}\theta_\kappa^M \times \left(s_{0\ell_i}^W\right)^\kappa$ from (C.15) are absorbed by the fixed effects. Since a direct interpretation of the individual parameters of regression (C.16) is not immediate, in Figure C.1 we summarize our estimation results by plotting the predicted gender gaps in standardized publications $\widehat{g}_W\left(s_{0\ell}^W\right) - \widehat{G}_M\left(s_{0\ell}^W\right) = \sum_{\kappa=0}^{3}\widehat{\gamma}_\kappa \times \left(s_{0\ell}^W\right)^\kappa$.

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



Notes: The Figure plots the estimated gender gap in standardized publications as a function of the percentage of female scientists by country and time period. Each circle corresponds to the estimated gender gap for the percentage of female scientists in a country in one of two periods: pre-WWII and post-WWII, e.g. United States - Pre WWII or United States - Post WWII. Gender gaps in standardized publications are computed from the estimates of equation (C.16). Countries with at least five women per cohort are included in the sample.

11 Further Details on the Predicted Citations Model

As outlined in the main text, we aim to account flexibly for the topic of each paper, which could have an influence on the citations of the paper. We therefore propose a ridge regression model that uses the words (unigrams) and word pairs (bigrams) that appear in the title of the 607,183 scientific papers we match to at least one scientist in the *Minerva* data. The model learns about finely grained fields of research and how many citations, on average, papers on a certain detailed topic receive.[58]

To prepare the data for the ridge regression model, 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 in the data and the size of the vocabulary V is the total number of unique unigrams and bigrams.[59]

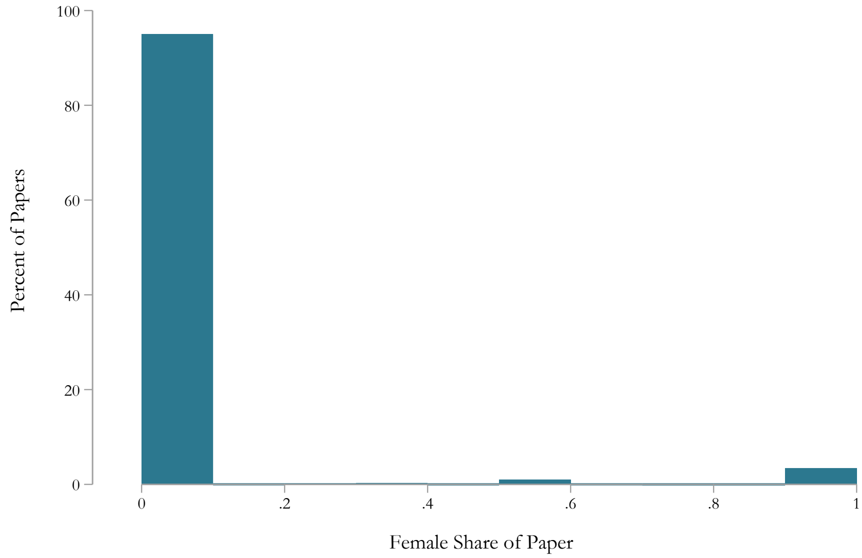
The model minimizes the loss function in equation 4 to identify the n-grams that have the highest predictive power for citations. The regularization term λ reduces overfitting of the model to the training sample, by picking up individual n-grams that appear in some extremely successful papers. We choose the optimal normalization strength using 10-fold cross-validation. To incorporate differences in citations for papers published in different time periods and disciplines, we fit the model separately for each of our cohorts and discipline. The model can thus account for the changing importance of topics over time and across disciplines.

If women (a) wrote papers with different titles than men and (b) were systematically under-cited because of discrimination, the model could internalize discrimination against papers written by women if we trained it on all papers (irrespective of the gender of the authors). To avoid such systematic bias, we train the model on the 81% papers which were written by only male authors (see Figure D.1 for a histogram of the female share of the papers in the analysis). We then use the estimated coefficients of equation 4 to predict citations for *all* papers (also those published by women).

[58]In independent research Hill and Stein (2021) also train a machine learning model to predict citations of academic research. The main difference to our approach, which uses the words in the title of papers, is that Hill and Stein (2021) use information from theProtein Data Bank in their prediction.

[59]We do not include trigrams in this model as they do not improve the out-of-sample performance and lead to overfitting.

Figure D.1: Histogram Female Share of Paper



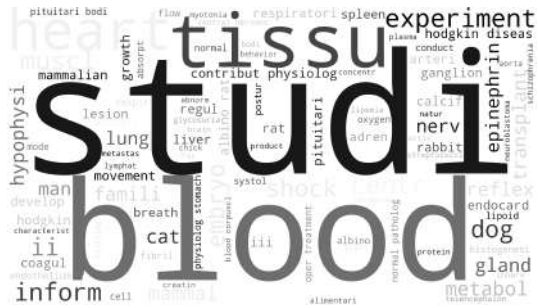
Notes: The Figure shows a histogram of the female share of papers for which we assign the gender for at least on of the authors.

11.1 Word Clouds Predicted Citations

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

(a) 1900

(b) 1914



(c) 1925

(d) 1938

(e) 1956

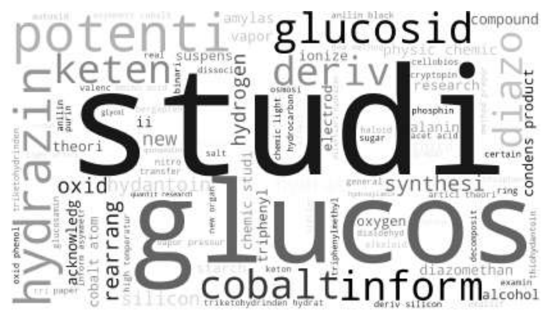
(f) 1969

Notes: The Figure shows the unigrams and bigrams with the largest predictive power for citations in medicine for each of the Minerva 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 5.1 for details.

Figure D.3: Words that Predict High Citations in Chemistry over Time

(a) 1900

(b) 1914

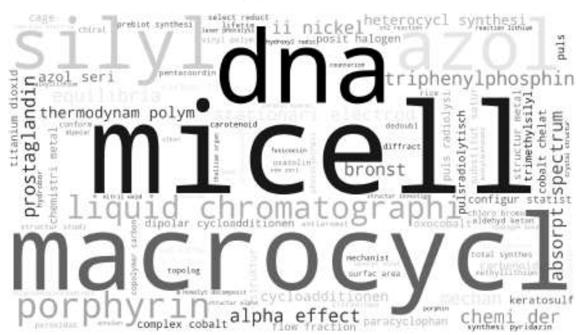


(c) 1925

(d) 1938

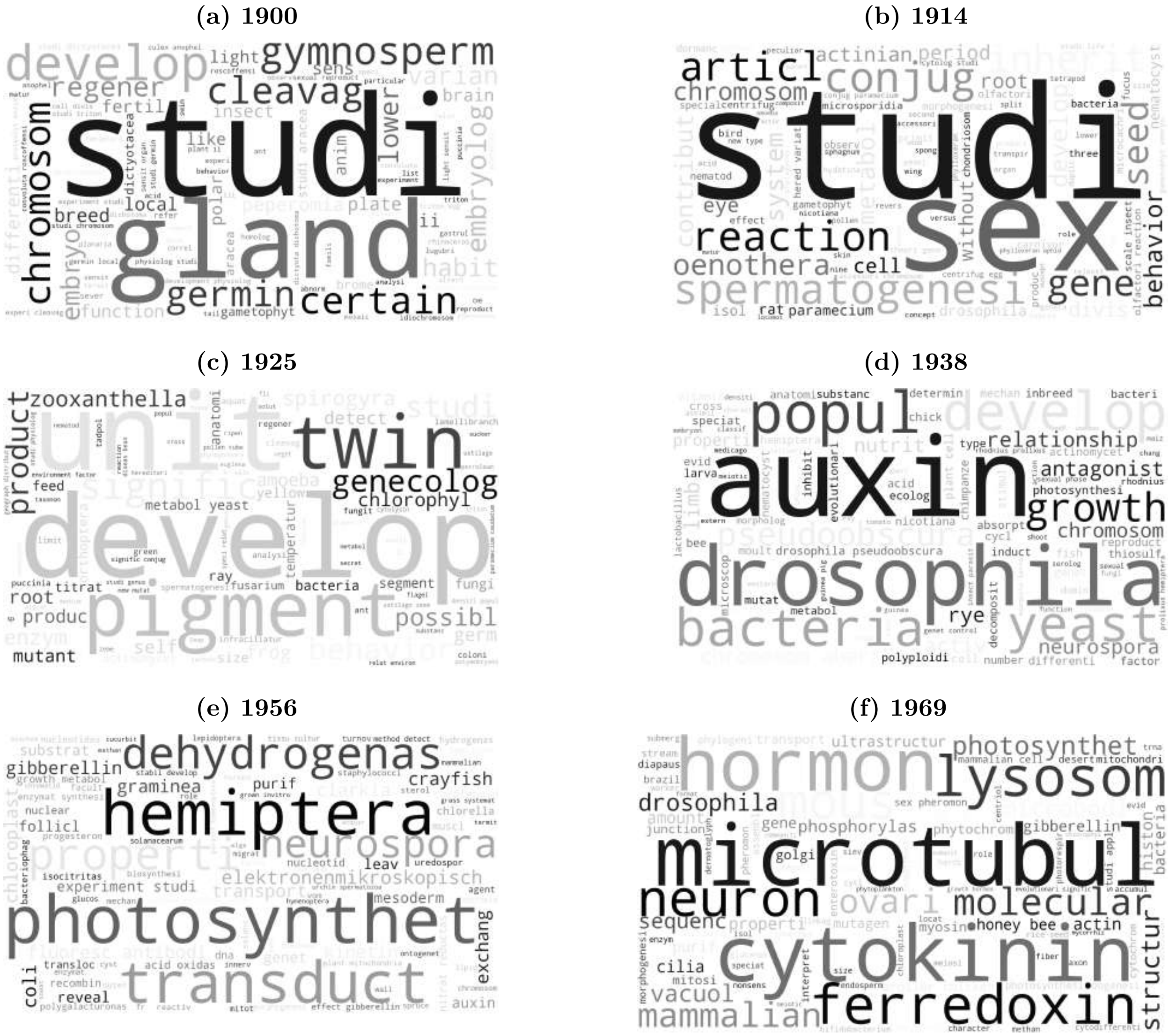
(e) 1956

(f) 1969



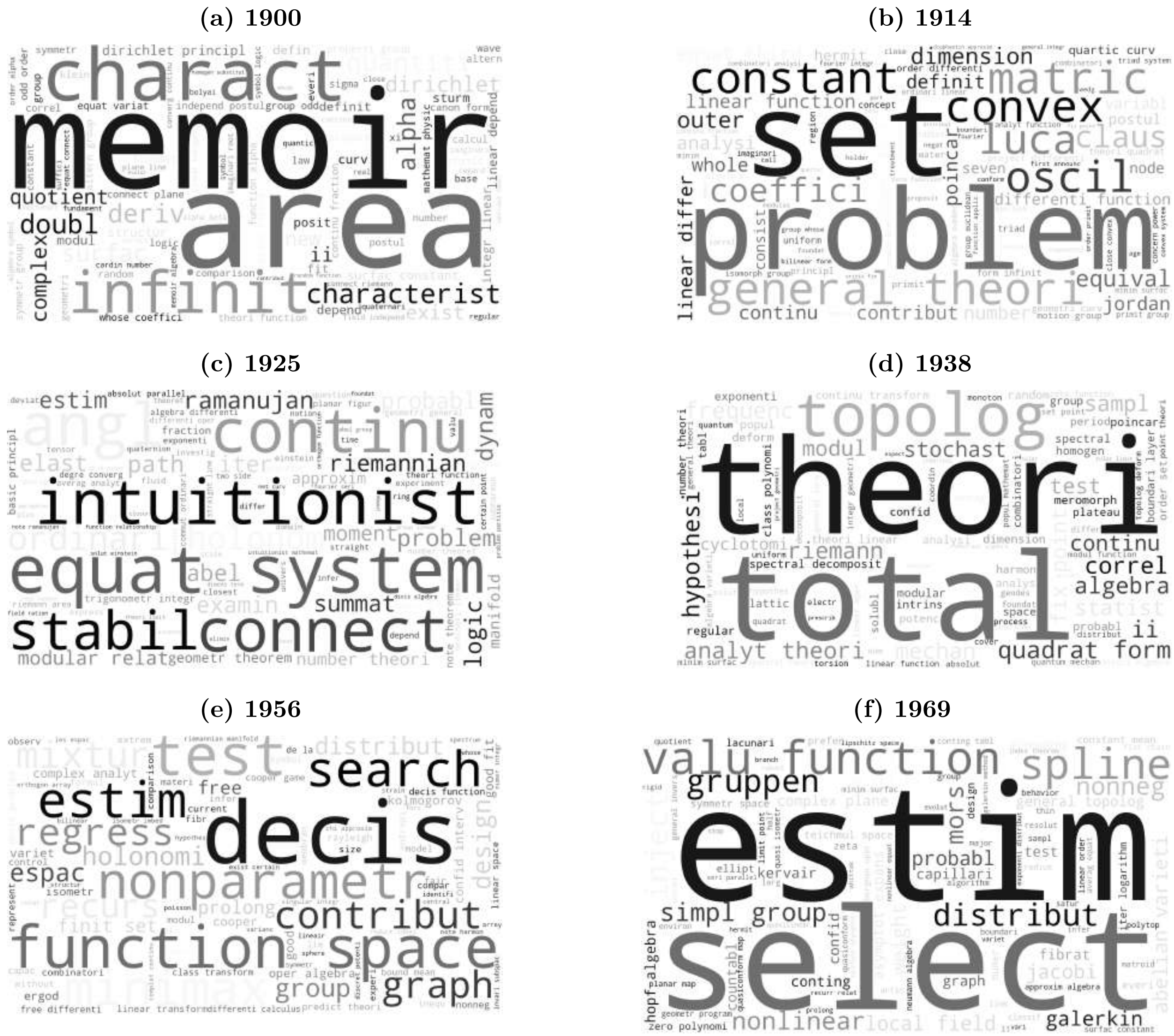
Notes: The Figure shows the unigrams and bigrams with the largest predictive power for citations in chemistry for each of the Minerva cohorts. The n-grams are identified with an L2-regularized regression model (ridge regression) that uses unigrams and bigrams of the title as inputs, see section 5.1 for details.

Figure D.4: Words that Predict High Citations in Biology over Time



Notes: The Figure shows the unigrams and bigrams with the largest predictive power for citations in biology for each of the Minerva 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 5.1 for details.

Figure D.5: Words that Predict High Citations in Mathematics over Time



Notes: The Figure shows the unigrams and bigrams with the largest predictive power for citations in mathematics for each of the Minerva 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 5.1 for details.

11.2 Censoring in the Citation Measure

Because of extreme outliers in citations, the ridge regression performs better if we transform p's total number of citations into permilles of the citation distribution within a discipline, country, and cohort (e.g. physics in the United States in 1900). One concern with transforming predicted citations into permilles is the presence of censoring in the dependent variable y_p, which measures the "position" of the citations of paper p in the distribution of citations across all papers in the same country, discipline, and cohort as p. As usual when working with citation data, a significant fraction of the papers does not receive any citation (12%

on average in our case). For any paper receiving 0 citations, we cannot determine the exact "position" in the distribution of citations. For example, if 12% of papers receive 0 citations, then all we know is that any paper with 0 citations lies somewhere in the interval 0-12% of the distribution of citations.OLS estimators can be inconsistent in the presence of censoring. Following standard results from the literature on partial identification (Manski, 2009), we can bound the extent of the censoring problem as we know that a paper with 0 citations has to be located somewhere between 0 and the quantile of papers with at least 1 citation. Suppose we observe a sample of three variables (y, x, d), with d an indicator denoting censoring: when $d = 0$, we observe x but not y. In our case, $d = 0$ for all those papers with 0 citations. The conditional expectation of y can then be expressed as:

$$\mathbb{E}\left[y|\, x\right] = \mathbb{E}\left[y|\, x, d=1\right] \times \Pr\left[d=1|\, x\right] + \mathbb{E}\left[y|\, x, d=0\right] \times \Pr\left[d=0|\, x\right].$$

The problem of censoring arises when $\Pr\left[d=0|\, x\right] > 0$. Without further assumptions, we can non-paramterically identify $\mathbb{E}\left[y|\, x, d=1\right]$ and $\Pr\left[d=1|\, x\right]$, but we cannot learn anything about $\mathbb{E}\left[y|\, x, d=0\right]$, and consequently about $\mathbb{E}\left[y|\, x\right]$, unless we have additional information to restrict the otherwise unbounded support of $\mathbb{E}\left[y|\, x, d=0\right]$. In our case, we observe natural bounds on $\mathbb{E}\left[y|\, x, d=0\right]$ since:

$$\underline{y}\left(x\right) \;\leq\; \mathbb{E}\left[y|\, x, d=0\right] \;\leq\; \overline{y}\left(x\right), \tag{D.1}$$

given by the appropriate quantiles of the citation distribution conditional on x, e.g., $\underline{y}\left(x\right)$ $= 0$ and $\overline{y}\left(x\right) = 120$ when y is measured in permilles (1000 quantiles) and there are 12% of papers with 0 citations.[60] This restriction in (D.1) then implies:

$$\begin{aligned} \mathbb{E}\left[y|\, x, d=1\right] \times \Pr\left[d=1|\, x\right] + \underline{y}\left(x\right) \times \Pr\left[d=0|\, x\right] \;\leq\; \mathbb{E}\left[y|\, x\right] \;\leq \\ \mathbb{E}\left[y|\, x, d=1\right] \times \Pr\left[d=1|\, x\right] + \overline{y}\left(x\right) \times \Pr\left[d=0|\, x\right]. \end{aligned} \tag{D.2}$$

Along these lines, to test the robustness of our estimates to the censoring in y_p, we repeat the predicted citation analysis assigning all the papers with 0 citations first to their lower bound $\underline{y}\left(x\right)$ and then to their upper bound $\overline{y}\left(x\right)$. Differently, in our baseline estimates we assign all papers with 0 citations to the midpoint $\frac{\overline{y}(x)-\underline{y}(x)}{2}$. The results from this exercise are

[60] Even though on average in our data 12 percent of papers have on average 0 citations, the share of papers with 0 citations can be higher or lower for any country-discipline-cohort cell used to compute the dependent variable y_p.

presented in Table D.1. Reassuringly, the exact allocation of the permilles to papers with 0 citations has no qualitative impact on our estimates and main findings.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles
Panel A: Baseline Permilles										
Share of Female Authors	-27.180***	-27.866***	-17.783***	-22.095***	-22.932***	-19.204***	-19.117***	-18.567***	-18.841***	-19.376***
	(4.035)	(4.096)	(5.362)	(3.802)	(4.237)	(3.064)	(3.140)	(2.540)	(2.603)	(2.569)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830
R^2	0.002	0.000	0.039	0.052	0.071	0.508	0.509	0.518	0.520	0.523
Panel B: Max Permilles										
Share of Female Authors	-29.053***	-27.055***	-17.180***	-21.376***	-22.079***	-18.374***	-19.189***	-18.601***	-19.115***	-19.598***
	(3.602)	(3.881)	(5.052)	(3.555)	(3.905)	(3.093)	(3.218)	(2.576)	(2.643)	(2.580)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830
R^2	0.016	0.019	0.059	0.072	0.089	0.514	0.517	0.526	0.529	0.531
Panel C: Min Permilles										
Share of Female Authors	-25.332***	-28.684***	-18.395***	-22.818***	-23.788***	-20.745***	-19.927***	-19.247***	-19.336***	-19.885***
	(5.238)	(4.332)	(5.679)	(4.062)	(4.579)	(3.089)	(3.144)	(2.791)	(2.795)	(2.775)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830
R^2	0.011	0.013	0.049	0.062	0.084	0.488	0.491	0.499	0.501	0.505

Notes: The Table shows gender gaps in citations. Results are estimated at the paper level. The dependent variable equals the permille in the cohort-country-discipline citation distribution of a paper. The main explanatory variable is the share of female authors of the paper. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for 1000 indicators that equal 1 if the paper falls into a certain permille of the predicted citation distribution. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model (ridge regression), see section 5.1 for details. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

Table D.2: Paper-Level Citations Gaps: Controlling for Predicted Citations - Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles	Citations Permilles
Panel A: Citation as Outcome										
Share of Female Authors	-4.933***	-5.268***	-2.847***	-3.292***	-3.374***	-3.044***	-3.276***	-2.142***	-2.161***	-2.235***
	(0.986)	(0.833)	(0.560)	(0.509)	(0.557)	(0.844)	(0.688)	(0.516)	(0.571)	(0.599)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830	518,830
R^2	0.009	0.011	0.020	0.022	0.020	0.059	0.060	0.065	0.066	0.063
Panel B: Gender of all Authors known										
Share of Female Authors	-14.331	-12.452	-4.823	-12.281	-16.426	-25.903***	-25.924***	-23.647***	-25.177***	-26.529***
	(9.648)	(9.108)	(12.852)	(11.051)	(10.967)	(4.129)	(3.949)	(3.866)	(3.821)	(3.775)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	138,207	138,207	138,207	138,207	138,207	138,207	138,207	138,207	138,207	138,207
R^2	0.028	0.032	0.060	0.077	0.108	0.519	0.522	0.527	0.530	0.536
Panel C: Out-of-Sample										
Share of Female Authors	-27.267***	-27.974***	-18.361***	-22.898***	-24.174***	-51.731	-55.943	-41.944	-51.590	-49.876
	(4.008)	(4.097)	(5.333)	(3.614)	(3.747)	(40.052)	(40.419)	(64.482)	(64.991)	(68.062)
Cohort FE	Yes					Yes				
Country FE	Yes					Yes				
Discipline FE	Yes					Yes				
Country × Discipline × Cohort FE		Yes	Yes	Yes			Yes	Yes	Yes	
University FE			Yes					Yes		
Department FE				Yes					Yes	
Cohort × Department FE					Yes					Yes
Predicted Citations (1000 quantile) FE						Yes	Yes	Yes	Yes	Yes
Observations	420,247	420,247	420,247	420,247	420,247	420,247	420,247	420,247	420,247	420,247
R^2	0.002	0.000	0.038	0.051	0.070	0.199	0.189	0.216	0.222	0.277

Notes: The Table shows gender gaps in citations. Results are estimated at the paper level. The dependent variable equals the permille in the cohort-country-discipline citation distribution of a paper. The main explanatory variable is the share of female authors of the paper. The regressions also control for various fixed effects, as indicated in the table. Additionally, the regressions control for 1000 indicators that equal 1 if the paper falls into a certain permille of the predicted citation distribution. Predicted citations are based on unigrams and bigrams of papers and estimated with a L2-regularized regression model (ridge regression), see section 5.1 for details. Standard errors are clustered at the discipline-country level. Significance levels: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

References

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

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

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

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

Hill, Ryan and Carolyn Stein. 2021. “Race to the bottom: Competition and quality in science.” *Working Paper* .

Manski, Charles F. 2009. *Identification for prediction and decision.* Harvard University Press.

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

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

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

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