Sources and Consequences of Systemic Content Bias: Evidence from Wikipedia

Nicole Venus*

July 29, 2025

Abstract

This paper examines the systemic bias against female scholars in Wikipedia content. Based on Wikipedia meta-data matched with rich panel data on a large sample of economists, psychologists and mathematicians, I estimate the gender gap in the likelihood that scholars receive a biographical entry conditioning on Wikipedia's own metric of relevance, the *notability criteria*. I show that while female economists are unconditionally around half as likely to have a biographical entry than their male colleagues, the gender gap in representation reduces to 9% conditioning on the notability criteria. This gap is even larger in psychology and reversed in mathematics. Over time the conditional gender gap in the representation of economists has closed, supported by Wikipedia editors organized in grassroots activist groups aimed at combating systemic bias on the platform. Leveraging the staggered introduction of a new content translation tool across language editions to predict page creations, I estimate the causal effect of having a biographical entry on Wikipedia on a researcher's news mentions. My findings underscore the importance of systemic biases: having a Wikipedia biography generates two additional news mentions per year. This demonstrates that content biases on digital knowledge platforms have implications that extend far beyond the platforms themselves, affecting which scientific knowledge is transmitted to a wider audience.

Keywords: user-generated content, digital platforms, economics profession, media bias
JEL codes: A14, H41, I23, J16, L86

*Institute of Economics (IdEP), Università della Svizzera Italiana (USI Lugano), Switzerland; email: nicole.venus@usi.ch
I am very grateful to Patricia Funk and Lorenz Küng for their academic guidance and continuous feedback on this project. I thank Aline Bütikofer, Nagore Iriberri, Fabrizio Colella, Raphael Parchet, Fabrizio Mazzonna, Felix Schönenberger, Eugenia Frezza and participants of the USI brown bag seminar, GRAPE Gender Gaps Conference, Young Swiss Economist Meeting, Gender and Economics Workshop Luxembourg, Annual Congress of the Swiss Society for Economics and Statistics, Universidad Pablo de Olavide research seminar in Sevilla, Gerzensee Alumni Conference, Digital Economy Workshop Berlin, Media Bias Workshop Frankfurt and Gender and Family Economics Workshop Paris for their valuable comments and suggestions. I acknowledge financial support from the Swiss National Science Foundation (grant no. 178887).

1 Introduction

Wikipedia is the most popular reference work in the world (The Economist, 2023) and expected to remain influential, as its content accounts for 3-5% of the training data which many modern LLMs, such as Chat-GPT, were trained on (Gertner, 2023). Since Wikipedia is a popular first point of reference for researchers, students, and journalists (Okoli et al., 2014), it has the potential to affect the visibility of researchers and their work on a large scale. Given its central role in the digital transmission of knowledge, allegations of systemic content bias on Wikipedia raised by both consumers and contributors are therefore deeply concerning. In particular, Wikipedia has been criticised for its substantial gender imbalance, as merely 15% of Wikipedia's contributors identify as female (Wikimedia, 2022) and less than 20% of biographical entries in the English-language edition of Wikipedia cover women (Humaniki Alpha, 2023). The case of Donna Strickland – a female physicist, who, at the time of being awarded the Nobel Prize, did not have a biographical entry on Wikipedia – sparked public debates on whether women in science were less likely to be represented than their equally notable male colleagues. In parallel, editors started organizing in activism groups, such as Women in Red, to fight a perceived gender bias in Wikipedia content.

Yet, there is a lack of empirical evidence on whether female scientists are less likely to have biographical entries because they are on average less notable, or whether they remain underrepresented, *even when accounting for their notability.* To quantify the gender gap in the likelihood that scholars receive a biographical entry conditional on their notability, I use Wikipedia's own metric of relevance – the so-called *notability criteria*. The main focus of this paper is on scholars in economics, a field in which women have historically been underrepresented, and despite some progress, continue to constitute the minority with a share of around 20% (Lundberg and Stearns, 2019). Furthermore, I delve into several mechanisms that have the potential to exacerbate or mitigate the systemic bias against female scholars on Wikipedia, such as grassroots activism on the platform. Since little is known about the consequences that such bias has for researchers themselves, I use the staggered introduction of a new content translation tool across language editions to estimate the causal effect of the representation of scholars on Wikipedia on their news mentions. The findings demonstrate that the systemic bias, present on digital knowledge platforms such as Wikipedia, propagates far beyond the platform itself, shaping which scientific knowledge is communicated to the general public.

This paper makes three main contributions. First, it provides the first estimate of the gender bias in the representation of economists on Wikipedia while rigourously conditioning on author notability. Wikipedia has defined a set of guidelines to gauge whether a person warrants a biographical entry, the so-called *notability criteria*. In contrast to other fields such as arts or politics, the notability of academics is determined by characteristics that are relatively easy to measure such as the number of citations, editorial positions held at prestigious academic journals, or important academic awards. Biographies on scholars who do not fulfill any of these criteria, are at risk of being removed from Wikipedia.

To quantify the gender bias in representation conditional on author notability, I collect Wikipedia meta-data through API requests and web-scraping from Wikipedia and match them with rich panel data of author characteristics for over 32,000 actively publishing economists spanning the years from 2001 to 2019 (Funk, Iriberri and Venus, 2024). I find that while female economists were unconditionally 53% less likely to have a Wikipedia entry than their male colleagues over the whole sample period, the gender gap in representation reduces to around 9% when conditioning on notability. This stems from a lack of page creations covering female economists, rather than the deletion of such pages. A comparison with the fields of psychology and mathematics – which differ substantially in their female share of the active research population (50% and 10%, respectively) – reveals that female scholars in more male-dominated fields are more likely to be represented on Wikipedia, conditional on notability. This is consistent with the hypothesis that women in these fields receive more recognition due to their higher visibility and organizations' incentives to signal diversity (Kanter, 1977).

Second, I show that the gender gap in representation in economics has closed over time due to contributions by editors affiliated with activism groups with a clear goal of fighting systemic bias against women in Wikipedia content. Furthermore, this paper also provides tentative evidence in support of the hypothesis that the conditional gender gap in the representation of economists is related to gender differences in using Wikipedia as a tool to self-promote.

Third, this paper highlights the importance of representation on Wikipedia for scholars by presenting the first causal evidence on the effect of having a Wikipedia page on the number of mentions a researcher receives in the news. I introduce a novel instrument to predict page creations by utilizing the staggered rollout of a new content translation tool across the six largest language editions in my sample, and in the second stage, estimate the effect on news mentions measured based on data collected from Google News. My findings show that being represented on Wikipedia significantly increases a researcher's news mentions by an average of two additional mentions per year, corresponding to 0.2 standard deviations, thereby boosting the visibility of their work outside academia.

Related literature. This paper adds to the growing evidence on gender differences in the representation or recognition of researchers in the field of economics: in the evaluation of their research (Abrevaya and Hamermesh, 2012; Card et al., 2020; Hengel, 2022; Hospido and Sanz, 2021; Iaria, Schwarz and Waldinger, 2022) and in the recognition of joint work (Sarsons, 2017; Sarsons et al., 2021); in applications (Eberhardt, Facchini and Rueda, 2023; Casarico and Rizzica, 2022) and promotions in academia (Bagues, Sylos-Labini and Zinovyeva, 2017), and in central banking (Hospido, Laeven and Lamo, 2022); in the selection probability into prestigious academic societies (Card et al., 2022, 2023) and editorial boards (Funk, Iriberri and Venus, 2024); in economics textbooks (Stevenson and Zlotnick, 2018); in research seminars (Dupas et al., 2021); and in posts on an anonymous online forum popular among economists (Wu, 2020). Compared to previous publications in the literature, this paper focuses on the role played by Wikipedia editors, often non-academic volunteers who decide on whom to represent and how, and on the readership of Wikipedia articles, which is not confined to members of the academic community. Hence, both in terms of who contributes to the gender gap in representation and who perceives it, this analysis focuses on the general public rather than solely on peers.

The discussion of potential mechanisms relates to prior research documenting the effectiveness of top-down affirmative action in reducing gender biases in peer recognition (Card et al., 2022). My findings show that community-driven initiatives serve as an effective alternative to top-down approaches. This paper also contributes to the growing literature on gender differences in self-promotion. Experimental evidence suggests that women are less likely to engage in self-promoting activities (Babcock et al., 2017; Exley and Kessler, 2022). Similar findings emerge from observational data in academia. For instance, female researchers are less likely to cite their own work (King et al., 2017) or promote their research on Twitter (Peng et al., 2025 forthcoming).

The paper also contributes to the relatively sparse literature examining gender differences in representation on Wikipedia.[1] Previous studies have either compared the coverage of Wikipedia

[1] The computer science literature refers to this type of bias as *coverage bias*. Other types of gender biases studied in the literature (Wagner et al., 2016) include gender differences in (i) the topics discussed e.g. biographies on women more frequently discuss the person's marital status (*topical biases*) (ii) the vocabulary used to describe a person e.g. biographies on women/men contain more negative/positive abstract terms, reflecting more

with an allegedly unbiased list of notable persons (Greenstein and Zhu, 2018; Reagle and Rhue, 2011), or attempted to find proxies for notability (e.g. Google search volumes; see Wagner et al., 2016). To my knowledge, Adams, Brückner and Naslund (2019) is the only published paper so far trying to estimate the gender gap in the representation of academics on Wikipedia conditional on proxies for notability.[2] Their analysis is confined to sociology faculty members in the US. Consistent with the findings presented in this paper, they find a significant unconditional and conditional under-representation of female sociologists based on data for the year 2016. However, their data set is cross-sectional, containing only few author characteristics. In contrast, this paper uses a rich panel data set, which allows conditioning on additional notability criteria such as fellowships, prizes and editorial positions, and to analyze the development of the conditional gender gap and its potential drivers over time.

Finally, this paper contributes to the meager body of literature studying the impact of Wikipedia content (Hinnosaar et al., 2023; Xu and Zhang, 2013). While social media presence has been shown to impact the credibility (Alabrese, Capozza and Garg, 2024), visibility and career success (Qiu et al., 2024) of researchers, the only paper examining the implications of Wikipedia content, I am aware of, is Thompson and Hanley (2018): They conduct an experiment, randomly selecting Wikipedia pages on science topics for upload, to estimate the effect on content and citation count of subsequently published peer-reviewed articles. However, so far no paper has attempted to investigate the potential consequences of misrepresentation outside academia – such as, in this paper, the impact on the visibility of scientists and their research in news media.

Overview. The rest of the paper is structured as follows. Section 2 gives an overview of Wikipedia's editing and deletion policies that are relevant for the analysis. Section 3 describes the data. Section 4 documents the unconditional gender gap. Section 5 discusses estimates of the conditional gender gap in representation and selection and provides a comparison between economics, psychology and mathematics. Section 6 shows that alternative reasons for notability do not drive baseline results and discusses structural biases i.e. gender differences in page visibility. Section 7 examines the role of initiatives promoting gender equality, page deletions, and gender differences in self-promotion. Section 8 provides causal evidence on the impact of representation on news mentions. Section 9 concludes.

negative/positive stereotypical language (*linguistic biases*) (iii) the visibility and accessibility of a biographical page e.g. how many links lead to the page (*structural biases*)

[2]There are a few (working) papers that try to estimate conditional gender gaps in representation very losely e.g. Schellekens, Holstege and Yasseri (2019) who only condition on the h-index.

2 Background

Wikipedia was founded in 2001, originally designed as an online platform to edit articles which should then be peer-reviewed by experts and published on the free-content encyclopedia Nupedia. Both platforms were initially funded by the web-advertising company Bomis. Since 2003, Wikipedia is hosted and financed by the non-profit organization Wikimedia Foundation.

Editing and deletion policies. There are three main user groups relevant for my analysis (Wikipedia, 2023*c*): unregistered users, registered users and administrators. All users that are not logged into a Wikipedia account are classified as unregistered users. Unregistered users can read Wikipedia pages, edit unprotected pages and draft new pages. However, they do not have the right to directly publish a new Wikipedia article. Registered users have additional rights, once their account is confirmed.[3] For instance, registered users can move drafts to the *mainspace*, i.e. transform a draft into a published article.[4] Administrators are registered users with more extensive rights, including the right to delete articles or block users. All actions taken by registered users are linked to their username. Edits made by unregistered users are linked to their IP address. Hereinafter, the terms *editors* and *contributors* refer to both unregistered and registered users who contribute to Wikipedia.

Any registered user can nominate a Wikipedia page for deletion (Wikipedia, 2023*a*). In the deletion discussion, editors can argue why they support or object a deletion. Reasons to support a deletion include, e.g., lack of notability of the article's subject, lack of reliable sources, or copyright violations. If, after seven days, there is a consensus to delete the page, the administrator proceeds with the deletion. If there is no consensus, the page is usually kept. Articles that will clearly not survive a deletion discussion can also be fast-tracked for deletion by an administrator. Deleted articles are only visible to administrators, while the deletion discussion is archived and visible to all users.

Notability. The most common argument brought forward to support the deletion of a biography article on an academic is the lack of notability. Wikipedia (2023*b*) sets out a list of criteria which make an academic notable. These include having a significant impact on their field, being awarded a prestigious academic prize or honor, being accepted as a member or fel-

[3] Typically, confirmation follows after 4 days and at least 10 edits. An exception is e.g. that the user operates from a blocked IP address. In urgent cases, confirmation can be fast-tracked by an administrator.

[4] Note that before December 2005 all users were allowed to directly publish a new article in the mainspace.

low at a selective academic society or having held an editorial position at a high-impact journal. Wikipedia (2023*b*) explains that the notability criteria for academics "are sometimes summed up as the 'Average Professor Test': When judged against the average impact of a researcher in a given field, does this researcher stand out as clearly more notable or more accomplished?" It is important to highlight that these criteria are guidelines, not strict rules. The decision to delete an article is based on the judgment of the editors whether a researcher fails the average professor test or not.

Activism among editors. In recent years, there has been growing attention to the gender gap among Wikipedia editors and rising concerns about a potential under-representation of women as subjects of biographical articles. In response, a couple of initiatives have been launched to close the asserted gender gap in Wikipedia's content and to encourage women and members of other under-represented groups to contribute to Wikipedia. One such initiative is the "Women in Red" project, which is aimed at promoting gender equality both in content and among editors. Members of this project regularly conduct so-called "Edit-a-thons", training and editing sessions for women.[5] Another example of such initiatives are the educational projects of the "Wiki Education Foundation". Teachers can register a class project in which students are asked to contribute to Wikipedia's content, for instance to increase the number of biography articles on female researchers.

3 Data

My analysis requires two main data sources: First, I need detailed academic records to account for the notability of an author. Second, I require Wikipedia meta-data which I collected via API requests and web scraping.

Academic records. I obtain publication and citation histories for over 32,000 economists from a panel data set constructed for and described in more detail in Card et al. (2022). This data set contains around 350,000 author-year observations on the universe of actively publishing economists. In each year of their active career, the data set reports the authors' cumulative number of publications in each of 36 high-impact journals[6] as well as the cumulative number of citations of papers published in each of the top-5 journals. Economists enter the sample as

[5]For a self-description of the initiative see `https://de.wikipedia.org/wiki/Women_in_Red`.
[6]For the complete list of Economics journals considered see Appendix Table 1.A1.

they start publishing in one out of the 36 journals and remain up to 18 years after their last publication or upon their death. The data also indicates the author's gender and, for each year, whether the author was a fellow of the Econometric Society.

From the data set collected for a follow-up paper (Card et al., 2023), I obtain information on whether the author was awarded a fellowship of the American Academy of Arts and Science, the National Academy of Sciences, or the Alfred P. Sloan Foundation. Funk, Iriberri and Venus (2024) augmented the data set with variables indicating, for each year and journal, whether the author held a position as an editor or associate editor at any of the top-5 or three general interest journals EJ, JEEA and REStat. I complemented this data with information on whether the author has, up to that year, received the Nobel Prize, John Bates Clark Medal, or the Frisch Medal. To estimate the gender gaps for psychology and mathematics, I obtain a sample of scholars along with their publication records from Card et al. (2023).[7]

Wikipedia meta-data. For each economist, I check if the author has a biographical entry in the English-language edition of Wikipedia.[8] If so, I retrieve when and by which registered user the article was created.[9] For each page, I collect the page classifiers and the number links leading to the page. For each creator of these entries, I collect the projects the user is affiliated with and how many other articles the editor created.[10]

In addition, for each author I check if there was a Wikipedia page on that author nominated for deletion (independent of what the result of that deletion discussion was). In a first step, I search for each author in Wikipedia's database of "Articles for deletion" and collect the search results. My search requires that both the author's first and the last name must be contained in the title and that at least one out of five keywords must be contained in the text of the deletion discussion.[11] In a second step, I hand-check the text of these search results to confirm that the deletion discussion was in fact written about that author. Note that for deleted articles, I only

[7]Appendix Table 1.A2 reports the set of journals considered for psychology and mathematics.

[8]In Funk, Iriberri and Venus (2024), I observe for each economist who is represented on Wikidata, the associated Wikidata ID. These IDs were originally collected to confirm the author's gender and (potentially) year of death. Wikidata is the database connecting all Wiki-projects such as Wikipedia, Wikinews, etc. All authors with a Wikipedia page have a Wikidata ID, but only a fraction of those represented on Wikidata have a Wikipedia page.

[9]Note that the revision history on Wikipedia not only tracks edits in the main space, but also in the draft and user space. This means that if e.g. a page was created in the draft space and then moved to the main space, the page creation date refers to the date at which the page was set up in the draft space. Similarly, the page creator would be the editor who set up the page in the draft space, not the one who moved it into the main space.

[10]For editors, specifying their gender in their profile page is optional and the majority of users chooses not to (Bayer, 2015). Therefore, I do not use this information in my analysis.

[11]These keywords are "econom*", "scholar*", "finan*", "university", and "academics", where * indicate wildcards.

observe the text in which contributors discuss why the article should be deleted, but I cannot see the deleted article itself or any information related to it.[12] Hence, in the case of deletion, I can only confirm that an article on that person existed (based on the deletion discussions and the references included in them) and when it was deleted, but not when the article was created.

For each author I define two outcome variables: an indicator variable equal to 1 for all years equal or after the year of page creation and 0 otherwise, and an indicator equal to 1 in the year of page creation and 0 otherwise. Finally, I exclude gender-ambiguous names which only make up 4% of the total number of author names. The data set covers the period starting from 2001, the year in which Wikipedia was founded, until 2019, the year in which the publication data set ends. The resulting data set allows me to predict the likelihood of having respectively of obtaining a biographical entry conditional on the notability criteria using a rich set of author characteristics. Summary statistics on the full set of actively publishing economists and on the set of economists during the years of being (respectively the year of becoming) represented on Wikipedia are reported in Appendix Tables 1.A3 – 1.A5. Statistics on the Wikipedia editors in my sample are reported in Appendix Table 1.A6.

Google news mentions. For evaluating the effect of having a Wikipedia page on the number of mentions a researcher receives in the news, I collected for each author represented in the English-language edition, in which other languages the author has a page and when those pages were created. Then I retrieved for each author the number of mentions in Google News. Since the computational cost increases with the number of languages, I focused on the six largest language editions in my sample, other than English: German, French, Italian, Spanish, Arabic and Portuguese. For each language, I translated the search request `‘‘author name’’ + economics | economist` with Google translate and gathered all results mentioning the request.[13] This yields a large data set of daily information on whether each author with an English Wikipedia page, has a page in any of the other languages and the number of mentions in news media. In a final step, I aggregated these observations into ten quarters reaching from 2018-Q3, two quarters before the introduction of the new translation tool, to 2020-Q4, five quarters after the tool was introduced in the last language I consider.

[12] This information is only visible to administrators since it can include sensitive data.
[13] Note that for Arabic, I also transcribed the name into Arabic.

4 Descriptive Statistics of the Unconditional Gender Gap

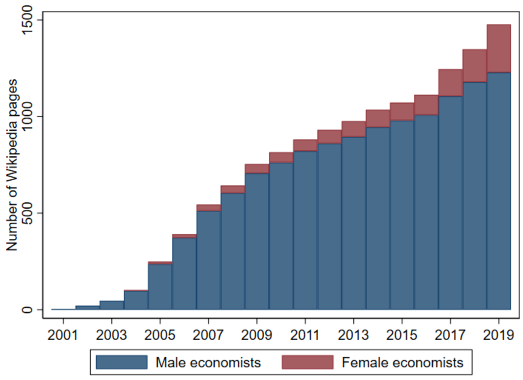
Figure 1 (a) shows the total number of female and male economists who are represented on Wikipedia. At the start of Wikipedia in 2001, only a handful of researchers had a page, whereas in 2019 almost 1,500 economists of my sample were represented on Wikipedia. The share of pages on female economists in total pages on all economists was steadily increasing over time and marked 16% in 2019 (panel (c), blue line). Before 2012, less than 10% of new biography articles on economists were created on female economists (panel (d), blue line). In the last four years of the sample, the share was well above 20%, reaching more than 50% in 2019.

To investigate the soaring share of new pages on female economists in recent years, I analyze the group affiliation of the Wikipedia editors who created the biography entries. To understand the role of initiatives promoting gender equality, I filter the groups users are affiliated with by their purpose. I identify three initiatives which specify closing Wikipedia's gender gap as their main goal: "Women in red", "Gender gap task force", and "Women scientists". I label all users belonging to at least one of those groups as "activist editors". In addition, I screen the class projects of those users who are classified as "Wiki Education student editors" for the purpose of the class and categorize only those who are affiliated to class projects with a clear gender focus as "activist editors".

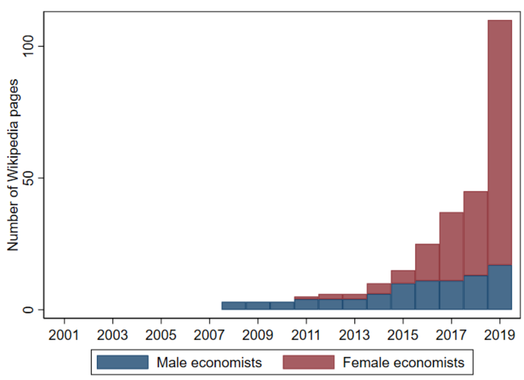
Figure 1 (b) shows the number of Wikipedia pages by activist editors. As expected, most of the articles they created until 2019 were on women, while pages on men are rare. Their impact is quite large. Users belonging to gender equality promoting initiatives created almost 100 new pages on female economists until 2019. As the total number of pages on female economists in the sample is around 250, they contributed around 40% of those pages. I recompute the shares of pages on female economists in total pages and in new pages excluding pages by affiliated editors (see red lines in panels (c) and (d)). Clearly, the sharp increase in the female share in new pages was due to contributions by affiliated editors, and when excluding their contributions, the female share in total pages would be at less than 11% in 2019.

Figure 1: Descriptive Statistics

Total number of Wikipedia pages

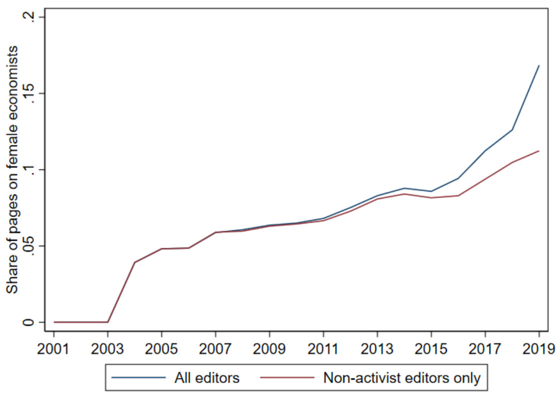


(a) By all editors

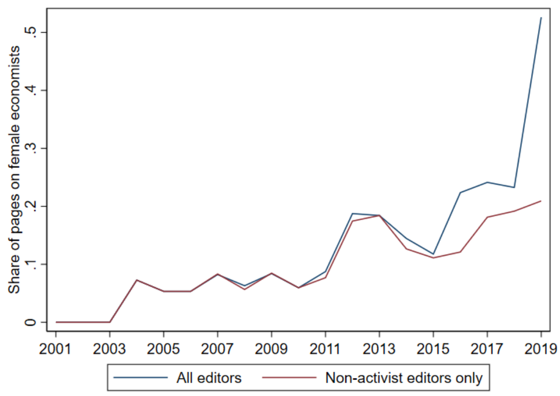


(b) By activist editors

Female share



(c) In total Wikipedia pages



(d) In new Wikipedia pages

The sample contains author-year observations for the universe of actively publishing economists of that year.

5 The Gender Gap Conditional on Author Characteristics

5.1 Representation

To estimate the gender gap in representation conditional on notability, I predict the probability of having a Wikipedia page controlling for gender and author characteristics. The outcome variable is an indicator equal to 1 in the years in which an economist is represented on Wikipedia and zero otherwise. Table 1 shows the average marginal effects estimated from a logistic regression. The predicted values are clustered around zero (see Appendix Figure 1.A1), which explains why the marginal effects deviate from the estimates of the linear probability model for some specifications (Appendix Table 1.A8). Given the distribution of the predicted values, the logistic regression model is preferred over a linear approximation. For all regressions, the corresponding estimates

Table 1: Representation: Baseline results, marginal effects

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.023	-0.025	-0.007	-0.004		
	(0.002)	(0.002)	(0.002)	(0.002)		
Female × (2001–2010)					-0.012	-0.010
					(0.003)	(0.003)
Female × (2011–2019)					-0.006	-0.003
					(0.002)	(0.002)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	350,314	350,314	350,314	350,314	350,314	350,314

The table shows the average marginal effects from a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

of the underlying latent variable model are reported in the Appendix.

Column 1 of Table 1 shows the unconditional gender gap in the representation of economists on Wikipedia, which is significantly negative. Unconditionally, female economists are 2.3 percentage points less likely to have a Wikipedia page than their male colleagues. Males in my sample have a baseline probability of 4.3% of having a page on Wikipedia (see Table 1.A3 row 2) so that female economists are 53% less likely to be represented than males. Controlling for year fixed effects in Column 2, the gender gap is still significantly negative.

Column 3 adds three sets of author characteristics. The first set of controls includes the cumulative number of publications in each of the 36 high-impact journals, non-parametric controls for the cumulative number of top-5 publications, and the inverse hyperbolic sine transformation (asinh from here on)[14] of the cumulative number of citations for papers published in each of the top-5 journals. Second, I control for whether the author held at least one position as editor or at least one position as associate editor at any of the top-5 and general interest journals in the respective year. Third, I include indicator variables equal to 1 if the author was a fellow of

[14]This transformation is used to approximate the natural logarithm while allowing for zeros. For $x > 2$, the $asinh(x)$ corresponds approximately to $\ln(2x)$.

the Econometric Society, the American Academy of Arts and Science, the National Academy of Sciences, or the Alfred P. Sloan Foundation, and if the author has received the Nobel Prize, the John Bates Clark Medal, or the Frisch Medal up to the respective year. Controlling for these author characteristics reduces the gender gap substantially, but does not eliminate it. It declines from 53% to 16% but remains statistically significant. Column 4 also conditions on the number of years since first publication to proxy for seniority. The conditional gender gap further decreases but remains marginally significant. Hence, in the most comprehensive specification, I find that female economists are conditionally 9% less likely to be represented on Wikipedia than males.

Columns 5 and 6 repeat the specifications of columns 3 and 4, but now interacting the female indicator with a decade indicator variable. The average effect over the entire sample period hides the fact that in recent years the gender gap is closing. In the 2000s, conditional on a wide set of author characteristics, women were strongly under-represented, while in the 2010s the gender gap was smaller (Column 5) or even became statistically indistinguishable from zero (Column 6) depending on the specification.

Figure 2 shows the development of the gender gap conditional on the same set of controls as in Column 6 of Table 1, over time. Clearly, female economists were conditionally under-represented until 2016.[15] In 2017 and 2018 the gap was not significantly different from zero anymore and in 2019 it even turned positive conditional on observables.

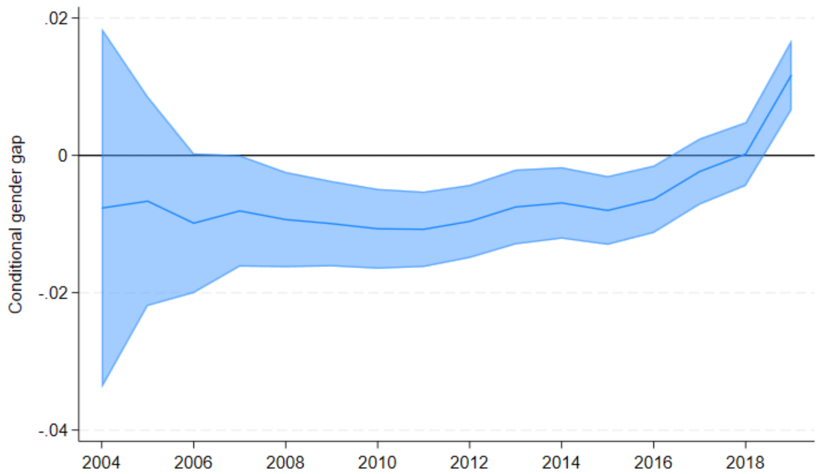
5.2 Selection

Table 2 repeats the regressions of Table 1 with exactly the same specifications, but under a different definition of the outcome variable. It focuses on the predictors of obtaining a Wikipedia page hence the outcome variable is now an indicator equal to 1 in the year in which an economist obtains a page on Wikipedia and zero otherwise. The sample is restricted to those who are *at risk* of getting a page, i.e. those who have not had a page up to the year before.[16] Column 1 shows that female economists are unconditionally significantly less likely to receive a page than their male colleagues. However, conditional on the same set of author characteristics as in Column 4 of Table 1, female academics actually had a significant advantage compared to males during the sample period. Looking at effect heterogeneity reveals that in the 2000s females still

[15]The conditional gender gap is not identified or very imprecisely estimated in the first few years of my sample since in those years no or only very few women were represented on Wikipedia.

[16]Note that this is a discrete-time approximation of the Cox hazard model (Efron, 1988).

Figure 2: Conditional gender gap over time



The figure shows the conditional gender gap (average marginal effects) obtained from specification (4) of the logistic regression in Table 1 (in the main text). The solid line shows the point estimates, the shaded area indicates the 95% confidence intervals.

Table 2: Selection: Baseline results, marginal effects

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.001	-0.001	0.002	0.002		
	(0.000)	(0.000)	(0.000)	(0.000)		
Female × (2001–2010)					-0.002	-0.002
					(0.000)	(0.000)
Female × (2011–2019)					0.004	0.005
					(0.001)	(0.001)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	338,443	338,443	338,443	338,443	338,443	338,443

The table shows the average marginal effects from a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist gets a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

had a significant disadvantage, while in recent years this effect has turned positive, with female economists being over-selected conditional on their characteristics.

5.3 Comparison across Disciplines: Psychology & Mathematics

Recent contributions to the literature suggest that the recognition of female scholars conditional on their academic performance is stronger in fields with a higher share of male scholars. For instance, women in male-dominated fields are more likely to be selected into reputable academic societies (Card et al., 2023) or to be appointed to assistance professorships (Belot, Kurmangaliyeva and Reuter, 2025). This pattern is consistent with the concept of tokenism (Kanter, 1977), whereby women in male-dominated fields receive more recognition conditional on their academic performance, due to higher visibility and organizational incentives to signal diversity. In order to examine whether similar patterns arise in the recognition by Wikipedia editors, who are predominantly non-academic volunteers, I compare the conditional gender gap in representation across three fields: economics, where women constitute around 20% of active scholars; psychology, with nearly 50% women; and mathematics, with only about 10% women.

As reported in the upper panel of Column 1 in Table 3, female psychologists are unconditionally 3 percentage points less likely to be represented on Wikipedia than their male peers. Given that males have a baseline probability of 4.9%, this translates to an unconditional gender gap of 61%. Conditioning on year fixed effects (Column 2) hardly changes the gender gap in representation. Column 3 conditions additionally on the cumulative number of publications in any of the psychology journals considered,[17] as well as on fellowship positions at the American Academy of Arts and Science and the National Academy of Sciences.[18] This reduces the gender gap conditional on notability criteria to 1.7 percentage points, or equivalently, a conditional gender gap of around 35%. When also controlling for the number of years since first publication (Column 4), the conditional gender gap reduces to 1.5 percentage points, which is significantly different from zero and translates to a gap of 31%. As shown in Column 5, the conditional gender gap, as measured in the widest specification decreased from 1.8 percentage points in the 2000s to 1.2 in the 2010s.

The unconditional gender gap in representation for math is positive but not significantly different from zero at the 5% significance level (see lower panel Column 1). Conditional on author notability (Column 3), women are significantly over-represented by 2.8 percentage points. Given the baseline probability of 4.3% for male mathematicians, women are, conditional on notability, 65% more likely to be represented on Wikipedia than their male colleagues. When additionally

[17] See Appendix Table 1.A2 for the full list of journals.

[18] Note that in contrast to scholars in economics, for psychology and mathematics I do not have information on editorial positions. In addition, these two subjects do not award a Nobel Prize.

accounting for academic age (Column 4), the gender gap increases to 3.1 percentage points, translating to an advantage for female mathematicians of 72%. Column 5 shows that while the gender gap estimated from the widest specification was essentially zero in the first decade, women were clearly over-represented in the 2010s.

Comparing the conditional gender gaps from the most comprehensive specification (Column 4) with the results for economics in Table 1, the gap is largest in psychology, followed by economics and reversed for mathematics, which aligns with the findings in Card et al. (2023) and Belot, Kurmangaliyeva and Reuter (2025). On the other hand, the dynamic over time is similar across the three disciplines.

Table 3: Representation: Psychologists and Mathematicians

	Psychology				
	(1)	(2)	(3)	(4)	(5)
Female	-0.030	-0.031	-0.017	-0.015	
	(0.002)	(0.002)	(0.002)	(0.002)	
Female × (2001–2010)					-0.018
					(0.002)
Female × (2011–2019)					-0.012
					(0.002)
Number of observations	382,040	382,040	382,040	382,040	382,040
	Mathematics				
	(1)	(2)	(3)	(4)	(5)
Female	0.007	0.004	0.028	0.031	
	(0.003)	(0.003)	(0.003)	(0.003)	
Female × (2001–2010)					-0.004
					(0.005)
Female × (2011–2019)					0.039
					(0.004)
Number of observations	381,673	381,673	381,673	381,673	381,673
Year fixed effects	no	yes	yes	yes	yes
Publications & fellowships	no	no	yes	yes	yes
Yrs since first publication	no	no	no	yes	yes

The table shows the average marginal effects from a logistic regression. The data set contains author-year observations for the universe of actively publishing psychologists (upper panel) and mathematicians (lower panel) of that year. The outcome variable is an indicator equal to 1 in the years in which a psychologist/mathematician has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

6 Robustness

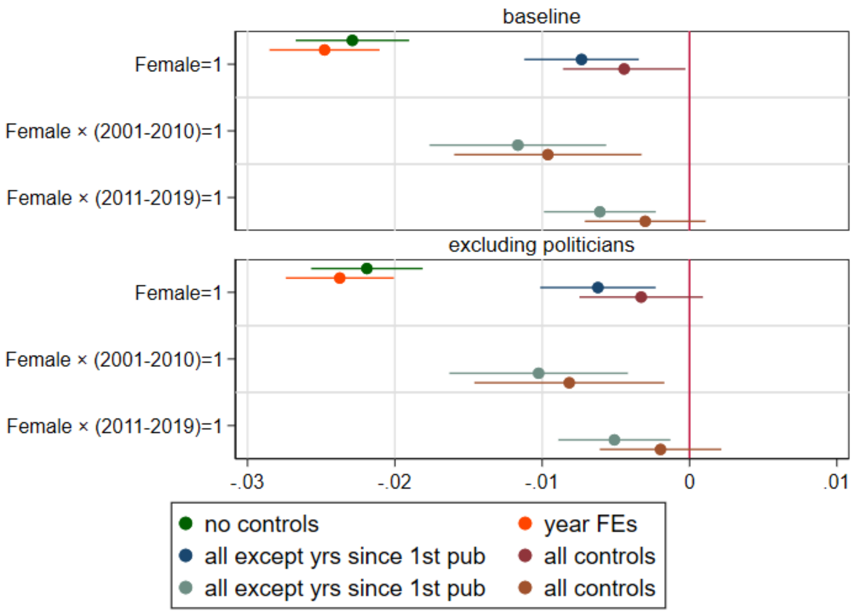
6.1 Alternative Reasons for Notability

One potential concern is that men could be more notable than women for other reasons than their achievements in the field of economics. Among other occupational categories in which economists fall, the most frequent one is "politician". Note that this definition not only encompasses "politicians" in the narrow sense such as presidents, chancellors, (prime) ministers or candidates running for office, but also high-ranking positions in international organizations and central banks, as well as leading advisory roles within governments.

If male economists are more likely to become notable as politicians than female, the conditional gender gap estimated above would be over-stated. A conservative approach to checking if this objection could drive the significant conditional gender gap, is to assume that all economists who also fall in the category of politicians are in fact notable for being a politician and not for being an economist.

To assess this possibility, I set the outcome variable to zero for all economists with a politician tag and re-estimate the baseline regression of Table 1. Figure 3 shows for all six specifications

Figure 3: Robustness Exercise



The figure shows the average marginal effects from the baseline regressions and the robustness exercise. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. For specifications, see text. Standard errors are clustered at the author-level. Dots denote point estimates and lines the 95% confidence interval.

in the upper panel the marginal effects from the baseline regressions and below the ones from the robustness exercise. Across specifications, the point estimates obtained from the robustness exercise are very close to the baseline estimates suggesting that gender differences in notability for being a politician do not drive the main results shown above.

6.2 Structural Biases

Gender differences in representation do not necessarily one-to-one translate into gender differences in visibility. If, for instance, pages on women are less often linked in other pages, then even if the gender gap in representation is closing over time, pages on female economists would still be less visible. Such a bias is called *structural bias* and is typically quantified by the difference in the number of links leading to a page.

To understand if structural biases are present in this case, I collect the number of links leading to each page in my sample. These links can be embedded in pages about other persons e.g. important co-authors of an economist, but also about a topic, e.g. a page on monetary theory referencing important scholars in the field.[19]

As shown in Table 4 column 1 pages on female economists are less likely to be referenced in other pages, i.e. less visible. However, once conditioning on the years since page creation (Column 2), there is no evidence for any structural bias. This means that pages on female economists are less visible on Wikipedia, but this is entirely related to the fact that they have been on the platform for a shorter time.

Table 4: OLS regression – Predictors of page links

	(1)	(2)
Female	-28.707	14.981
	(9.197)	(9.144)
Years since page creation		9.721
		(0.686)
Number of observations	1,416	1,416

The table shows the estimates from a linear regression. The data set contains all pages on economists from my sample that were available in 2019. The outcome variable, measured as of July 2024, is the number of links leading to a page.

[19]Note that due to data availability, I can only observe the current number of links leading to a page, but not obtain the history over time.

7 Potential Mechanisms

7.1 The Role of Page Deletions

A conditional gender gap in representation on Wikipedia can stem from two sources. On the one hand, editors could create pages on female economists less frequently than on equally notable male economists. On the other hand, even if male and female economists had the same conditional probability of having a page created in their name, more frequent deletion of pages on women compared to men could generate a gender gap. In the analysis above, the focus was on pages that were created and not deleted up to the point of data collection. To understand if the second channel is at play, I pool the authors who had a page with those authors whose page was deleted in each year. These authors together are those which, in that year, had a page and of which some pages were deleted and the others remained. Based on that sub-sample, I define two outcome variables, being nominated for deletion and having one's page deleted. Table 5 shows the results of a logistic regression for both dependent variables, once estimated without controls (Columns 1 and 3) and once controlling for author characteristics (Columns 2 and 4).

While pages on female economists have a higher probability of being nominated for deletion and deleted, only the unconditional gender gap in nominations is substantial and statistically significantly positive. These results are consistent with findings in the literature. Tripodi (2023) finds that biographies on women are (unconditionally) more likely to be nominated for deletion

Table 5: Page deletions: estimates

	Nominated for deletion		Page deleted	
	(1)	(2)	(3)	(4)
Female	0.755	0.277	0.896	0.302
	(0.278)	(0.280)	(0.565)	(0.577)
Author characteristics	no	yes	no	yes
Number of observations	13,676	13,676	13,676	13,676
Pseudo R-squared	0.01	0.08	0.01	0.18

The table shows the estimates from the latent model of a logistic regression. The data set contains author-year observations for the universe of actively publishing economists who had a page on the English-language Wikipedia that year. The outcome variable is an indicator equal to 1 in the year in which an economist's page is nominated for deletion or is deleted and zero otherwise. Author characteristics are the cumulative number of papers, cumulative number of top-5 publications, asinh of citations in each top-5 journal and years since first publication. Standard errors in parenthesis are clustered at the author-level.

while Adams, Brückner and Naslund (2019) show that pages on female academics are not significantly more likely to be deleted. However, two caveats have to be kept in mind. First, I do not observe when deleted pages had been created, i.e. they are missing in the sub-sample of pages that could potentially be deleted in the years before their deletion. Second, nominations for deletion, and even more so page deletions, are very rare events in my sample, so the estimates are very imprecise. At the same time, since these events are very rare, we can conclude that even if pages of female economists were more likely deleted conditional on author characteristics, page deletions are unlikely to drive much of the overall conditional gender gap.

7.2 Gender Differences in Self-Promotion

A potential reason for the under-representation of female economists could be gender differences in using Wikipedia as a tool to self-promote. If male economists are more prone to write autobiographies than women, this could explain why female economists are less likely to be represented on Wikipedia. The experimental literature has shown that women are less likely to self-promote, both with and without incentives to do so (Exley and Kessler, 2022). While I cannot directly observe whether page subjects created their own articles, some patterns suggest that such a mechanism might be at play. Editors who are merely interested in creating articles about themselves will see little benefit in creating other articles. Hence, if self-promotion differs by gender and self-promoting editors create only one page, then the share of pages on female economists would differ between editors who created only one page and editors who created more than one.

Table 6: Self-promotion: estimates

	(1)	(2)
>1 pages created	0.070	0.053
	(0.024)	(0.022)
Year fixed effects	No	Yes
Number of observations	1,014	1,014

The table shows the estimates from an OLS regression. The data set contains all editors of the pages on actively publishing economists within my sample. The unit of observation is the editor level. The dependent variable is the share of pages on female economists. The main explanatory variable is an indicator equal to 1 if the editor has created more than one page (within or outside my sample of economists) and zero otherwise. If an editor is not registered (i.e. only an IP address is visible), the independent variable is set to missing. For student editors (activist and non-activist), the independent variable is set to 1.

To test this hypothesis, I obtain for all editors who created an article on an economist in my sample the total number of pages they created.[20] As reported in the summary statistics on editors in Table 1.A6, on average around 61% of editors created more than one page. For all editors I compute the share of females in the total number of pages on economists they created and define an indicator variable equal to 1 if the editor has created more than one page in total. For student editors, I set the indicator variable equal to 1 since most of them only created one page but, due to their role, they are not suspect of creating a page for self-promotion. As shown in Table 6, among editors who have created more than one page or are student editors (i.e. who are not suspect of self-promotion), the share of pages on female economists is significantly higher than among those who created only one page (Column 1). Even when controlling for the year of the first contribution, editors in the non-suspect group are around 5 percentage points more likely to write on a female economist than those in the suspect group (Column 2). Using this proxy, this evidence supports the hypothesis that men are more prone to creating a biography in their own name.

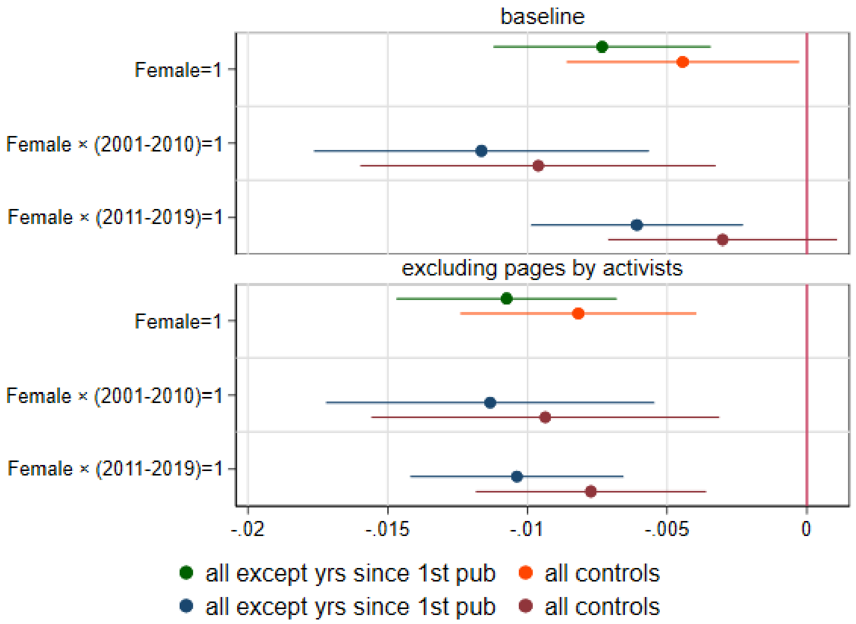
7.3 The Role of Initiatives promoting Gender Equality

In order to understand the role of activist editors in the closing (and reversal) of the conditional gender gap over time, I re-estimate the specifications of Table 1 but set the outcome variable to 0 for those economists whose page was set up by an editor from an activist group. Figure 4 shows the marginal effects from the baseline regressions in the upper panel and below the ones assuming that the pages created by activist editors would not have been created. The figure highlights that the conditional gender gap estimated in specifications 3 and 4 would indeed be considerably larger if pages created by activist editors would not have been created. In the most comprehensive specification, I find that the conditional gender gap would be 0.8 percentage points instead of 0.4 percentage points as in the baseline (Online Appendix Table 1.A11 Column 4). Specifications 5 and 6 show that this deviation stems from recent years, which is in line with Figure 1b documenting a large rise in the number of articles on female economists created by activist users in the most recent years of the sample.

Specifications 3 and 4 of Figure 5 show that if the pages created by activist editors would not have been created, the overall conditional gender gap in selection would be essentially zero. Hence, the significant advantage for female economists documented in the upper panel is in

[20] These pages can be on a person (within or outside the sample) or on any other topic.

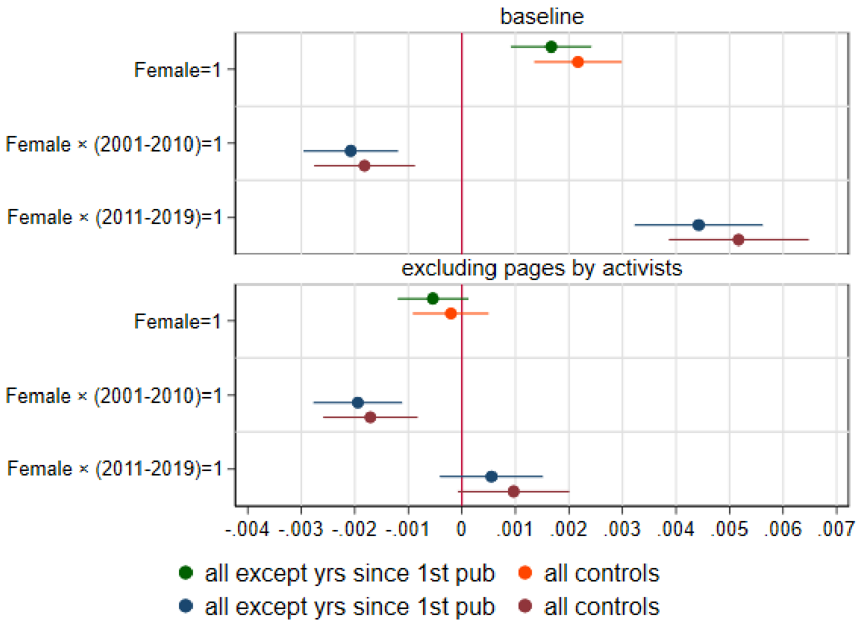
Figure 4: The role of initiatives promoting gender equality in representation



The figure shows the marginal effects from the baseline regressions and the mechanism exercise. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. For specifications, see text. Standard errors are clustered at the author-level. Dots denote point estimates and lines the 95% confidence interval.

fact due to activist users' work. If pages created by activists would not have been created, the conditional gender gap in selection would have been significantly negative in the 2000s and effectively zero from 2011 onwards (see specifications 5 and 6). In other words, in recent years, editors not affiliated with activist groups have selected male and female economists equally conditional on notability, while in the early years of Wikipedia they were significantly more likely to select men.

Figure 5: The role of initiatives promoting gender equality in selection



The figure showes the average marginal effects from the baseline regressions and the mechanism exercise. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist gets a page on the English-language Wikipedia and zero otherwise. For specifications, see text. Standard errors are clustered at the author-level. Dots denote point estimates and lines the 95% confidence interval.

8 The Effect of Representation on News Mentions

As discussed before, a misrepresentation on Wikipedia can have a range of potential consequences, both within academia and beyond. In this section, I focus on one particular implication: the effect on visibility in news media. Numerous descriptive studies, such as Okoli et al. (2014), have documented that journalists often rely on Wikipedia for collecting background information on a topic. This raises an important question: does the under-representation of female scientists on Wikipedia reduce their visibility in the media? For instance, under-representation could result in fewer references to their research in journalistic work or in a lower likelihood of being invited for expert interviews.

8.1 Identification

When estimating the causal effect of Wikipedia representation on media mentions, a key identification problem arises: unobserved factors may simultaneously cause a spike in news mentions and trigger the creation of a Wikipedia page for that particular scholar. For instance, a scholar's co-author receiving the Nobel Prize or the sudden surge of interest in a specific field in which a scholar specializes (e.g. the heightened attention to epidemiological models due to the onset of the Covid-19 pandemic) could trigger both a spike in media attention and the creation of a Wikipedia page for the scholar.

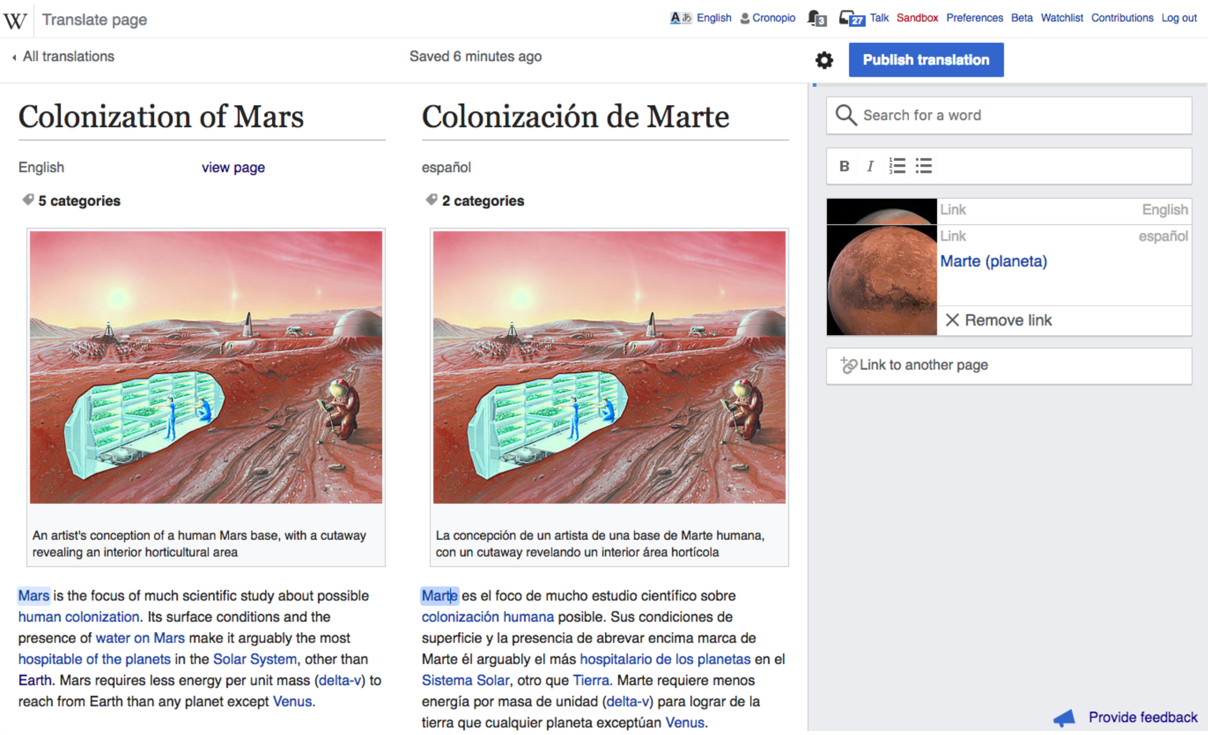
Instrument. To address this identification problem, I propose a novel instrument that leverages the staggered rollout of a new translation tool across different Wikipedia language editions. The tool that eases page translations between languages was introduced in several steps: After an initial period of development and testing, during which the tool was only available to a few selected Wikipedia editors, the Beta feature of the tool was released to all editors in 2015. A major revision occurred in 2019, marked by a shift from Apertim or Yandex to Google's Neural Machine Translation model as the underlying translation model, which substantially improved the quality of translations. In comparison to previous translation models, Google's transformer-based model is estimated to reduce translation errors between major language pairs by 55% to 85%.[21]

The integration of the new content translation tool into the Wikipedia editor interface was staggered across the six largest language editions in my sample: While Google translation from English to Arabic, French, Spanish and Portuguese became available in the first quarter of 2019, the tool was introduced for Italian in Q3 and for German in Q4. As explained in Zhu and Walker (2025), the timing of the rollout was due to differences in the length of the community governance processes across language editions.

As visualized in Figure 6, the tool is embedded in an interface which allows editors a side-by-side comparison between the original and the translation. The tool allows editors to generate a draft by machine translation, which they can then review and edit before publishing in the target language. In addition to translating the text, it also allows for automatic transfers of tables, graphs, links and other page features between language editions. It is important to stress that the content translation tool is not designed to fully automate article translations but to assist human editors by making the translation process more efficient. Indeed, Zhu and Walker (2025) show that the integration of Google's Neural Machine Translation model into Wikipedia's editor interface has significantly increased content production through translation.

[21] For more information on the performance see `https://research.google/blog/a-neural-network-for-machine-translation-at-production-scale/`

Figure 6: Content translation interface



Machine translation draft from English to Spanish.
Source: https://commons.wikimedia.org/wiki/File:Cx-screenshot-aug-2017.png

Identification strategy. My identification strategy is designed as follows: For all economists with a page in the English edition of Wikipedia in my sample, define instrument Z^j_{lt} indicating if the translation tool was available to translate a page from English into target language l at time t for j periods. This instrument is used to predict a variable indicating whether author i, who is already represented in the English-language edition, also has a page P_{ilt} in the target language edition l at time t. The predicted probability to have a page $\hat{P}_{ilt}$ is in turn used to estimate the effect of being represented in language edition l on the number of times the author i is mentioned in the news in language l in the same time period, N_{ilt}. I focus on the period from Q3 2018, two quarters before the tool was introduced in the first four languages I consider, until Q4 2020, five quarters after the tool was introduced in the last language. Since the page creation date and news mentions are daily data, I aggregate them by quarter.

Hence, the first-stage equation follows an event-study design:

$$P_{ilt} = \sum_{j \neq -1} \alpha_j Z^j_{lt} + \eta_i + \zeta_t + \xi_l + \beta X_{lt} + \varepsilon_{ilt}$$

where Z^j_{lt} denote the event-time dummies. The first-stage conditions on author, time and language fixed effects, as well as in some specifications, on language×time varying covariates X_{lt}. This specification implies that I allow the impact of the new tool on the probability to have a page in another language to vary across time.

In the second-stage, the effect of the predicted likelihood to have a page $\hat{P}_{ilt}$ on news mentions is estimated conditioning on the same author, time and language fixed effects and covariates X_{lt} as in the first stage

$$N_{ilt} = \gamma \hat{P}_{ilt} + \eta_i + \zeta_t + \xi_l + \delta X_{lt} + \varepsilon_{ilt}$$

Threats to the validity of the instrument. This identification scheme relies on the assumption that – conditional on language, time and author fixed effects - in the absence of the implementation of the content translation tool the outcome variable in earlier and later treated target languages would have evolved in parallel. As evidence in support of this assumption, I show event-time dummies of the first-stage below. First-stage results confirm that language editions treated earlier and later followed similar trends in page creations before the implementation of the tool. In addition, no anticipation effects are visible from the event-time dummies.

Any language-specific shocks to news reporting, which correlate with the timing of the tool implementation, would threaten my identification strategy. For instance, an economic downturn in a country, which makes up a large fraction of editors for a specific language edition and coincides with the implementation of the tool, could both affect editing activity and news about economics. As a robustness check to address this concern, I also estimate a specification controlling for the number of edits made in target language l in time t.

Another threat to my identification strategy would be that the content translation tool affects news mentions through other channels than page creations. For instance, the tool might not only affect the quantity of articles created, but also the quality thereof. To address this concern, I additionally estimate a specification controlling for the average quality of new articles,[22] measured as the ratio of deleted articles to total pages created, varying across languages and time periods.[23]

A potential concern would also be that the innovation not only affects the supply-side of Wikipedia content, but also the demand-side i.e. the ability of journalists to read articles in

[22]Note that another concern would be that the tool affects the news mentions through the quality of existing pages. A robustness check to analyze the validity of this concern would require information on the quality of edits (in contrast to article creations), which I leave for future work.

[23]I.e. For each quarter and language, I collected all articles created (these can be within or outside my sample), and calculated the proportion that have been deleted by the date of data collection (February 2025).

their main language. However, Google's Neural Machine Translation model was already available to users consuming web content, e.g. via Google Chrome's browser plug-in, before July 2018. So during my sample period, the demand-side of Wikipedia content was not affected by the innovation.

Finally, the implementation of the tool could also affect the demand for content. If the tool draws journalists' attention to Wikipedia (e.g. through local media reporting about the innovation), journalists might increase their general consumption of Wikipedia content and therefore be more likely to cite scholars with a biography on the platform. While I cannot fully rule out this mechanism, given the technical nature of the innovation, a strong direct effect of the rollout on news mentions due to increased awareness does not appear plausible.

8.2 Results

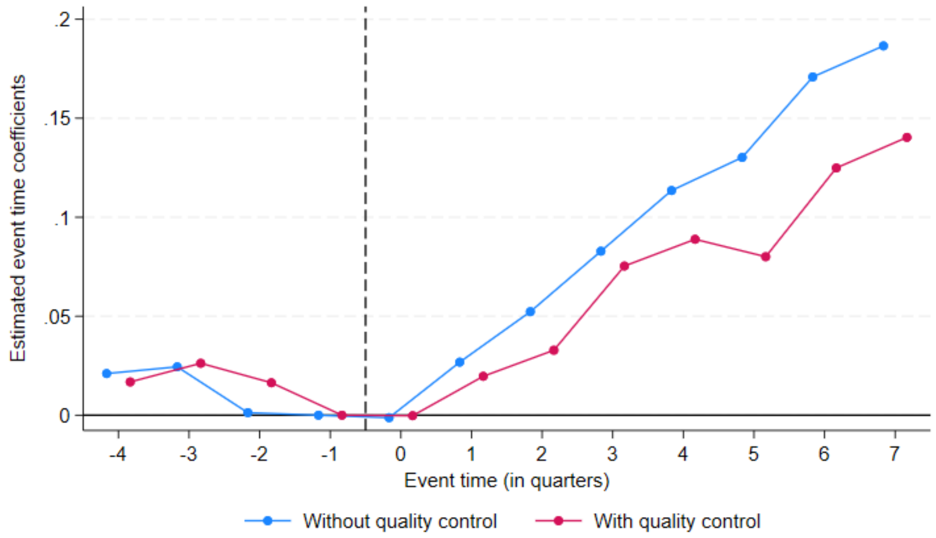
Figure 7 presents the estimated event-time coefficients from the first stage both with and without quality controls. The estimates are shown relative to the quarter before the introduction of the tool in $j = -1$. Reassuringly, the estimated event-time dummies before the introduction are stable and close to the coefficient in the last quarter before the introduction, suggesting that neither pre-trends nor anticipation effects are a concern in the first stage.

After the adoption, the likelihood to have a page in a treated target language starts to rise and ascends continuously over time. Two quarters after the introduction, the probability to have a page is around 6 percentage points higher compared to the last quarter pre-adoption, after five quarters around 10 and after eight quarters 20 percentage points higher (denoted in blue).

Note that the event-time dummies are consistently lower when conditioning on the deletion ratio (denoted in red), suggesting that page creations induced by the tool come, at least to some degree, at the cost of lower average page quality. Table 7 shows that across all specifications, the Kleibergen-Paap test for weak identification yields an F-statistic well above 10, suggesting a strong first stage.

In the second stage, I estimate the effect of the probability of having a page in language l, based on first-stage event time dummies, on news mentions in the same language. As shown in the main specification in Table 7 Column 1, having a Wikipedia page increases the number of times a researcher is mentioned in the news by 0.5 mentions per quarter. This effect is statistically significant at the 5% level, translating into approximately 2 additional mentions per year, which corresponds to around 0.2 standard deviations.

Figure 7: First stage – Event Time Coefficients



The figure shows the event time coefficients from the first stage regression relative to the quarter before the introduction of the content translation tool ($t = -1$). Regressions condition on time, author and language fixed effects (in blue) and average page quality, measured as the deletion ratio (in red).

In Column 2, I re-estimate the equation, additionally conditioning on the deletion ratio as a proxy for quality in target language l in time t. Conditioning on page quality, the estimated effect of representation on news mentions is even larger. This suggests that failing to account for the tool's impact on page quality does not drive the positive effect but rather exerts a slight downward bias on the coefficient of interest. The bias arises because the content translation tool reduces average page quality, and, all else equal, lower-quality pages receive fewer news mentions.

In Column 3 I also control for total number of edits made in a language edition. Even though, this takes away some of the variation induced by the innovation, the coefficient changes only slightly compared to the second specification. Finally, I repeat all three specifications using a Poisson model. As seen in Column 4 to 6, the results remain robust even when deviating from the linear specification.

Table 7: IV - Second stage

	Linear			Poisson		
	(1)	(2)	(3)	(4)	(5)	(6)
Page	0.505	0.656	0.574	1.934	2.590	3.162
	(0.217)	(0.310)	(0.357)	(0.934)	(1.188)	(1.555)
Language FE	yes	yes	yes	yes	yes	yes
Time FE	yes	yes	yes	yes	yes	yes
Author FE	yes	yes	yes	yes	yes	yes
Quality control	no	yes	yes	no	yes	yes
Number of edits	no	no	yes	no	no	yes
Number of observations	54300	54300	54300	43980	43980	43980
Kleibergen-Paap rk Wald F statistic	51.729	51.548	49.705	51.729	51.548	49.705

The table shows the effects of the predicted probability to have a page (based on the first stage) on the number of news mentions an author receives in the same language. The dataset contains author-language-quarter observations for all authors with an English Wikipedia page before 2018Q3. Standard errors in parentheses are clustered at author-level. The estimates from the Poisson model are computed following a control function approach, in which – in addition to the predicted probability of having a page – I also condition on the residuals from the first stage.

9 Conclusion

Based on a data set comprising over 32,000 active economists, this paper estimates the unconditional and conditional gender gaps in being and becoming represented on Wikipedia. Female economists are unconditionally 53% less likely than their male colleagues to be represented on Wikipedia. Controlling for author characteristics reduces the gender gap in representation to 9%. Additionally, I demonstrate that factors unrelated to academic achievement do not explain these results. I find no evidence for structural biases, i.e. pages on male and female economists are equally visible. A comparison across disciplines reveals that the gender gap conditional on notability is larger in Psychology than in Economics, and reversed in Mathematics.

The conditional gender gap in representation closed over time. This development was driven by two forces. Editors affiliated with initiatives promoting gender equality in the representation of academics on Wikipedia have significantly over-selected female economists conditional on author characteristics. At the same time, non-affiliated editors, who were conditionally under-selecting women in the 2000s, have selected their biographical subjects in a gender-neutral way in the 2010s. The conditional gender gap in representation is mainly driven by a lack of page creations on women and not by page deletions, and is related to gender differences in using Wikipedia as a tool to self-promote.

Furthermore, this paper underscores the importance of representation on Wikipedia for scholars by providing the first causal evidence on the effect of having a Wikipedia page on the number of mentions a researcher receives in the news. I propose a novel instrument to predict page creations leveraging the staggered introduction of a new content translation tool across language editions. My results demonstrate that representation on Wikipedia matters: having a page significantly increases the number of mentions researchers receive in the news and thereby boosts the visibility of their work in the public eye. As such, initiatives aimed at alleviating the under-representation of female scholars on Wikipedia not only enhance their visibility within the encyclopedia but also extend their impact to other domains, such as news media.

References

Abrevaya, Jason, and Daniel S Hamermesh. 2012. "Charity and favoritism in the field: Are female economists nicer (to each other)?" *Review of Economics and Statistics*, 94(1): 202–207.

Adams, Julia, Hannah Brückner, and Cambria Naslund. 2019. "Who counts as a notable sociologist on wikipedia? gender, race, and the "professor test"." *Socius*, 5: 1–14.

Alabrese, Eleonora, Francesco Capozza, and Prashant Garg. 2024. "Politicized scientists: Credibility cost of political expression on twitter." CESifo Working Paper No. 11254.

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): 714–747.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the gender composition of scientific committees matter?" *American Economic Review*, 107(4): 1207–1238.

Bayer, Tilman. 2015. "How many women edit Wikipedia?" Published April 30; https://wikimediafoundation.org/news/2015/04/30/how-many-women-edit-wikipedia/.

Belot, Michèle, Madina Kurmangaliyeva, and Johanna Luise Reuter. 2025. "Gender diversity and diversity of ideas."

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

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2022. "Gender differences in peer recognition by economists." *Econometrica*, 90(5): 1937–1971.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2023. "Gender gaps at the academies." *Proceedings of the National Academy of Sciences*, 120(4).

Casarico, Alessandra, and Lucia Rizzica. 2022. "Women in economics: the role of gendered references at entry in the profession." CEPR Discussion Paper 17474.

Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, et al. 2021. "Gender and the dynamics of economics seminars." National Bureau of Economic Research Working Paper 28494.

Eberhardt, Markus, Giovanni Facchini, and Valeria Rueda. 2023. "Gender differences in reference letters: Evidence from the economics job market." *The Economic Journal*, 133(655): 2676–2708.

Efron, Bradley. 1988. "Logistic regression, survival analysis, and the Kaplan-Meier curve." *Journal of the American Statistical Association*, 83(402): 414–425.

Exley, Christine L, and Judd B Kessler. 2022. "The gender gap in self-promotion." *The Quarterly Journal of Economics*, 137(3): 1345–1381.

Funk, Patricia, Nagore Iriberri, and Nicole Venus. 2024. "Women in Editorial Boards: An Investigation of Female Representation in Top Economic Journals." CEPR Discussion Paper 19303.

Gertner, Jon. 2023. "Wikipedia's moment of truth." *New York Times*. Published July 18; Updated September 8; https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html; Accessed on September 19, 2023.

Greenstein, Shane, and Feng Zhu. 2018. "Do experts or crowd-based models produce more bias? Evidence from Encyclopedia Britannica and Wikipedia." *Mis Quarterly*, 42(3): 945–960.

Hengel, Erin. 2022. "Publishing while female: Are women held to higher standards? Evidence from peer review." *The Economic Journal*, 132(648): 2951–2991.

Hinnosaar, Marit, Toomas Hinnosaar, Michael Kummer, and Olga Slivko. 2023. "Wikipedia matters." *Journal of Economics & Management Strategy*, 32(3): 657–669.

Hospido, Laura, and Carlos Sanz. 2021. "Gender gaps in the evaluation of research: evidence from submissions to economics conferences." *Oxford Bulletin of Economics and Statistics*, 83(3): 590–618.

Hospido, Laura, Luc Laeven, and Ana Lamo. 2022. "The gender promotion gap: evidence from central banking." *Review of Economics and Statistics*, 104(5): 981–996.

Humaniki Alpha. 2023. "Gender metrics." https://humaniki.wmcloud.org/search; Accessed on June 6, 2023.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2022. "Gender Gaps in Academia: Global Evidence Over the Twentieth Century." https://dx.doi.org/10.2139/ssrn.4150221; Accessed on September 24, 2023.

Kanter, Rosabeth Moss. 1977. *Men and Women of the Corporation.* New York:Basic Books.

King, Molly M, Carl T Bergstrom, Shelley J Correll, Jennifer Jacquet, and Jevin D West. 2017. "Men set their own cites high: Gender and self-citation across fields and over time." *Socius*, 3.

Lundberg, Shelly, and Jenna Stearns. 2019. "Women in economics: Stalled progress." *Journal of Economic Perspectives*, 33(1): 3–22.

Okoli, Chitu, Mohamad Mehdi, Mostafa Mesgari, Finn Årup Nielsen, and Arto Lanamäki. 2014. "Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership." *Journal of the Association for Information Science and Technology*, 65(12): 2381–2403.

Peng, Hao, Misha Teplitskiy, Daniel M Romero, and Emőke-Ágnes Horvát. 2025 forthcoming. "The Gender Gap in Scholarly Self-Promotion on Social Media." *Nature Communications.* Working paper version at https://arxiv.org/abs/2206.05330.

Qiu, Jingyi, Yan Chen, Alain Cohn, and Alvin E Roth. 2024. "Social media and job market success: A field experiment on Twitter." *Available at SSRN 4778120.*

Reagle, Joseph, and Lauren Rhue. 2011. "Gender bias in Wikipedia and Britannica." *International Journal of Communication*, 5: 1138–1158.

Sarsons, Heather. 2017. "Recognition for group work: Gender differences in academia." *American Economic Review*, 107(5): 141–145.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2021. "Gender differences in recognition for group work." *Journal of Political Economy*, 129(1): 101–147.

Schellekens, Menno H, Floris Holstege, and Taha Yasseri. 2019. "Female scholars need to achieve more for equal public recognition." *arXiv preprint arXiv:1904.06310.*

Stevenson, Betsey, and Hanna Zlotnick. 2018. "Representations of men and women in introductory economics textbooks." *AEA Papers and Proceedings*, 108: 180–185.

The Economist. 2023. "Wikipedia is 20, and its reputation has never been higher." Published in the January 9th 2021 edition under the headline 'The other tech giant'; https://www.economist.com/international/2021/01/09/wikipedia-is-20-and-its-reputation-has-never-been-higher; Accessed on December 31, 2023.

Thompson, Neil, and Douglas Hanley. 2018. "Science is shaped by Wikipedia: evidence from a randomized control trial." MIT Sloan Research Paper 5238-17.

Tripodi, Francesca. 2023. "Ms. Categorized: Gender, notability, and inequality on Wikipedia." *New Media & Society*, 25(7): 1687–1707.

Wagner, Claudia, Eduardo Graells-Garrido, David Garcia, and Filippo Menczer. 2016. "Women through the glass ceiling: gender asymmetries in Wikipedia." *EPJ Data Science*, 5: 1–24.

Wikimedia. 2022. "Community Insights 2021 Report." https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2021_Report; Accessed on June 6, 2023.

Wikipedia. 2023*a*. "Wikipedia:Deletion policy." https://en.wikipedia.org/w/index.php?title=Wikipedia:Deletion_policy&oldid=1170037569; Accessed on June 13, 2023.

Wikipedia. 2023*b*. "Wikipedia:Notability (academics)." https://en.wikipedia.org/w/index.php?title=Wikipedia:Notability_(academics)&oldid=1168415952; Accessed on June 13, 2023.

Wikipedia. 2023*c*. "Wikipedia:User groups." https://en.wikipedia.org/w/index.php?title=Wikipedia:User_access_levels&oldid=1168558307#Autoconfirmed_and_confirmed_users; Accessed on June 13, 2023.

Wu, Alice H. 2020. "Gender bias among professionals: an identity-based interpretation." *Review of Economics and Statistics*, 102(5): 867–880.

Xu, Sean Xin, and Xiaoquan Zhang. 2013. "Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction." *Mis Quarterly*, 1043–1068.

Zhu, Kai, and Dylan Walker. 2025. "Machine-assisted Content Creation on Peer Production Platforms." *Available at SSRN.*

1.A Additional Figures and Tables

Table 1.A1: List of Journals: Economics

A. Top-5 journals	
American Economic Review	Quarterly Journal of Economics
Econometrica	Review of Economic Studies
Journal of Political Economy	
B. General interest journals	
Economic Journal	Review of Economics and Statistics
Journal of European Economic Association	
C. Selected top-field journals	
Journal of Development Economics	Journal of Labor Economics
Journal of Econometrics	Journal of Monetary Economics
Journal of Economic Theory	Journal of Public Economics
Journal of Finance	
D. Other journals	
American Economic Journal: Applied Economics	International Journal of Game Theory
American Economic Journal: Economic Policy	Journal of American Statistical Association
American Economic Journal: Macroeconomics	Journal of Economic History
American Economic Journal: Microeconomics	Journal of Economic Literature
American Economic Review: Papers and Proceedings	Journal of Economic Perspectives
Econometric Theory	Journal of Health Economics
Economic Theory	Journal of International Economics
Economica	Journal of Mathematical Economics
Games and Economic Behavior	Quantitative Economics
International Economic Review	Rand Journal of Economics
Theoretical Economics	

Notes: The table lists all 36 journals included in the dataset of the actively publishing economists in Card et al. (2022).

Table 1.A2: List of Journals: Psychology & Mathematics

Psychology Journals	**Mathematics Journals**
Journal Of Personality And Social Psychology	Annals Of Statistics
Psychological Review	Inventiones Mathematicae
Cognition	Proceedings Of National Academy Of Sciences
Child Development	Journal Of American Mathematical Society
Cognitive Psychology	Duke Mathematical Journal
American Psychologist	Journal Of American Statistical Association
Psychological Science	Journal Of Computational Physics
Psychological Bulletin	Acta Mathematica
Trends In Cognitive Sciences	Annals Of Probability
Annual Review Of Psychology	Transactions Of American Mathematical Society
Journal Of Experimental Psychology: General	Annals Of Mathematical Statistics
Proceedings Of National Academy Of Sciences	American Journal Of Mathematics
Developmental Psychology	Advances In Mathematics

Notes: The table lists all journals included in the dataset in Card et al. (2023).

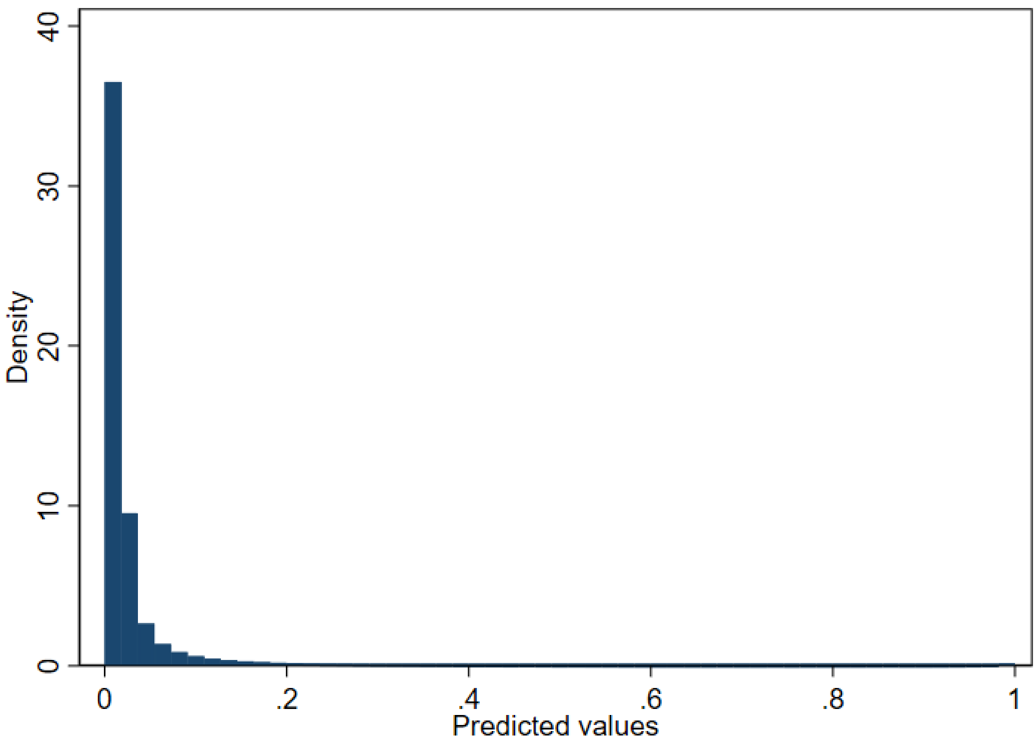


Figure 1.A1: Predicted values from logistic regression (representation)

The figure shows the density of the predicted values obtained from specification (6) of the logistic regression in Table 1.A7.

Table 1.A3: Summary statistics for actively publishing economists

	2001-2019			2001-2010			2011-2019		
	All	Male	Female	All	Male	Female	All	Male	Female
Female	0.179	0.000	1.000	0.154	0.000	1.000	0.199	0.000	1.000
Has Wikipedia page	0.039	0.043	0.020	0.023	0.025	0.009	0.052	0.058	0.027
Years since first publication	12.880	13.663	9.287	12.861	13.568	8.990	12.894	13.743	9.473
A. Cum. publications in:									
Econometrica	0.221	0.258	0.051	0.242	0.276	0.053	0.204	0.242	0.050
REStud	0.145	0.166	0.052	0.150	0.169	0.049	0.141	0.163	0.054
AER	0.271	0.300	0.140	0.275	0.301	0.135	0.268	0.299	0.143
QJE	0.147	0.163	0.070	0.157	0.172	0.071	0.139	0.156	0.069
JPE	0.168	0.192	0.060	0.195	0.217	0.072	0.147	0.171	0.052
B. Asinh cum. citations in:									
Econometrica	0.439	0.503	0.144	0.450	0.507	0.136	0.430	0.500	0.149
REStud	0.313	0.350	0.144	0.290	0.321	0.115	0.332	0.375	0.161
QJE	0.376	0.412	0.210	0.366	0.399	0.189	0.383	0.423	0.224
AER	0.658	0.710	0.419	0.626	0.670	0.383	0.684	0.745	0.441
JPE	0.427	0.477	0.197	0.456	0.500	0.217	0.403	0.457	0.184
C. Editorial positions:									
Editor	0.003	0.003	0.002	0.003	0.003	0.001	0.003	0.003	0.002
Associate editor	0.013	0.014	0.011	0.014	0.015	0.012	0.013	0.013	0.011
D. Fellowships and prizes:									
EMA	0.032	0.037	0.007	0.035	0.040	0.007	0.030	0.035	0.008
AAAS	0.014	0.016	0.004	0.016	0.018	0.004	0.013	0.015	0.004
NAS	0.004	0.004	0.001	0.004	0.005	0.001	0.004	0.004	0.001
AEAF	0.003	0.004	0.001	0.004	0.004	0.001	0.003	0.004	0.001
Sloan	0.012	0.013	0.010	0.012	0.013	0.011	0.012	0.013	0.009
Nobel prize	0.002	0.002	0.000	0.002	0.003	0.000	0.002	0.002	0.000
Clark medal	0.001	0.002	0.001	0.002	0.002	0.000	0.001	0.001	0.001
Frisch medal	0.001	0.002	0.000	0.001	0.002	0.000	0.001	0.002	0.000
Number of authors	49,513	40,143	9,370	21,314	17,778	3,536	28,199	22,365	5,834

Table 1.A4: Summary statistics for economists who are represented on Wikipedia

	2001-2019			2001-2010			2011-2019		
	All	Male	Female	All	Male	Female	All	Male	Female
Female	0.093	0.000	1.000	0.058	0.000	1.000	0.105	0.000	1.000
Years since first publication	27.973	28.693	20.919	27.747	28.046	22.889	28.053	28.935	20.534
A. Cum. publications in:									
Econometrica	1.313	1.422	0.240	1.459	1.533	0.246	1.261	1.381	0.238
REStud	0.812	0.870	0.240	0.897	0.938	0.237	0.782	0.845	0.241
AER	1.569	1.637	0.904	1.687	1.747	0.705	1.527	1.596	0.942
QJE	1.163	1.192	0.882	1.256	1.282	0.841	1.131	1.159	0.890
JPE	1.136	1.206	0.450	1.378	1.428	0.560	1.051	1.124	0.428
B. Asinh cum. citations in:									
Econometrica	2.053	2.194	0.673	2.095	2.185	0.638	2.039	2.198	0.680
REStud	1.545	1.630	0.708	1.502	1.561	0.541	1.560	1.656	0.741
QJE	2.396	2.411	2.251	2.241	2.266	1.845	2.451	2.465	2.330
AER	3.026	3.075	2.548	2.955	2.994	2.318	3.051	3.105	2.593
JPE	2.412	2.513	1.419	2.538	2.610	1.358	2.367	2.477	1.430
C. Editorial positions:									
Editor	0.020	0.019	0.032	0.022	0.022	0.019	0.020	0.018	0.034
Associate editor	0.038	0.034	0.070	0.041	0.040	0.048	0.036	0.032	0.075
D. Fellowships and prizes:									
EMA	0.303	0.320	0.142	0.348	0.361	0.145	0.287	0.304	0.141
AAAS	0.228	0.236	0.146	0.305	0.310	0.222	0.200	0.209	0.131
NAS	0.069	0.071	0.052	0.101	0.101	0.092	0.058	0.060	0.044
AEAF	0.062	0.064	0.044	0.090	0.090	0.101	0.052	0.055	0.033
Sloan	0.097	0.094	0.130	0.086	0.087	0.077	0.101	0.096	0.141
Nobel prize	0.043	0.047	0.004	0.076	0.080	0.010	0.032	0.035	0.003
Clark medal	0.032	0.033	0.025	0.051	0.053	0.024	0.026	0.026	0.026
Frisch medal	0.016	0.018	0.000	0.017	0.018	0.000	0.016	0.017	0.000
Number of authors	2,615	2,277	338	888	832	56	1,727	1,445	282

Table 1.A5: Summary statistics for economists who become represented on Wikipedia

	2001-2019			2001-2010			2011-2019		
	All	Male	Female	All	Male	Female	All	Male	Female
Female	0.156	0.000	1.000	0.063	0.000	1.000	0.246	0.000	1.000
Years since first publication	23.745	24.664	18.765	25.120	25.471	19.873	22.410	23.690	18.491
A. Cum. publications in:									
Econometrica	1.000	1.132	0.285	1.300	1.373	0.218	0.708	0.841	0.302
REStud	0.626	0.699	0.231	0.772	0.812	0.164	0.484	0.562	0.248
AER	1.216	1.282	0.859	1.378	1.421	0.727	1.059	1.113	0.892
QJE	0.866	0.910	0.625	1.058	1.077	0.782	0.678	0.709	0.586
JPE	0.864	0.958	0.354	1.126	1.167	0.509	0.610	0.706	0.315
B. Asinh cum. citations in:									
Econometrica	1.624	1.781	0.777	1.864	1.952	0.556	1.392	1.574	0.832
REStud	1.270	1.368	0.737	1.324	1.386	0.409	1.217	1.347	0.819
QJE	1.958	1.974	1.871	1.982	1.993	1.807	1.934	1.950	1.886
AER	2.534	2.561	2.385	2.551	2.569	2.280	2.517	2.551	2.411
JPE	1.878	2.014	1.142	2.150	2.216	1.161	1.614	1.770	1.137
C. Editorial positions:									
Editor	0.024	0.021	0.040	0.023	0.023	0.018	0.024	0.018	0.045
Associate editor	0.065	0.055	0.116	0.054	0.052	0.073	0.075	0.059	0.126
D. Fellowships and prizes:									
EMA	0.226	0.249	0.101	0.297	0.309	0.109	0.156	0.175	0.099
AAAS	0.142	0.158	0.058	0.231	0.238	0.127	0.057	0.062	0.041
NAS	0.037	0.040	0.018	0.067	0.068	0.055	0.007	0.006	0.009
AEAF	0.031	0.034	0.018	0.061	0.060	0.073	0.003	0.003	0.005
Sloan	0.088	0.083	0.116	0.084	0.084	0.091	0.091	0.081	0.122
Nobel prize	0.020	0.024	0.000	0.041	0.044	0.000	0.000	0.000	0.000
Clark medal	0.018	0.020	0.007	0.032	0.033	0.018	0.004	0.004	0.005
Frisch medal	0.011	0.013	0.000	0.013	0.013	0.000	0.009	0.012	0.000
Number of authors	1,778	1,501	277	876	821	55	902	680	222

Table 1.A6: Summary statistics on editors

	All	Not identified	Student editor	Other editors
More than 1 page	0.607		0.043	0.634
Share of pages on female economists	0.165	0.058	1.000	0.124
# of pages in sample	1.618		1.000	1.648
Total # of pages	361.578		1.043	379.101
Year of first page creation	2,010.633	2,006.540	2,019.000	2,010.226
# of editors	1,127	113	47	967

Table 1.A7: Representation: Baseline results, estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.782	-0.869	-0.307	-0.182		
	(0.083)	(0.084)	(0.088)	(0.090)		
Female × (2001–2010)					-0.539	-0.433
					(0.169)	(0.168)
Female × (2011–2019)					-0.253	-0.122
					(0.085)	(0.087)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	350,314	350,314	350,314	350,314	350,314	350,314

The table shows the estimates from the latent model of a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

Table 1.A8: Representation: Baseline results, linear probability model

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.023	-0.026	-0.005	-0.003		
	(0.002)	(0.002)	(0.002)	(0.002)		
Female × (2001-2010)					0.002	0.004
					(0.002)	(0.001)
Female × (2011-2019)					-0.010	-0.008
					(0.002)	(0.002)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	350,314	350,314	350,314	350,314	350,314	350,314

The table shows the estimates from a linear probability model. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

Table 1.A9: Selection: Baseline results, marginal effects

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.190	-0.204	0.311	0.393		
	(0.064)	(0.063)	(0.066)	(0.067)		
Female × (2001–2010)					-0.527	-0.447
					(0.143)	(0.143)
Female × (2011–2019)					0.702	0.790
					(0.077)	(0.079)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	338,443	338,443	338,443	338,443	338,443	338,443

The table shows the estimates from the latent model of a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist gets a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

Table 1.A10: Representation: Robustness exercise, estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.775	-0.863	-0.267	-0.138		
	(0.085)	(0.085)	(0.091)	(0.093)		
Female × (2001–2010)					-0.482	-0.374
					(0.171)	(0.170)
Female × (2011–2019)					-0.217	-0.082
					(0.087)	(0.089)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	350,314	350,314	350,314	350,314	350,314	350,314

The table shows the estimates from the latent model of a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

Table 1.A11: Representation: Mechanism exercise, marginal effects

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.025	-0.027	-0.011	-0.008		
	(0.002)	(0.002)	(0.002)	(0.002)		
Female × (2001–2010)					-0.011	-0.009
					(0.003)	(0.003)
Female × (2011–2019)					-0.010	-0.008
					(0.002)	(0.002)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	350,314	350,314	350,314	350,314	350,314	350,314

The table shows the average marginal effects from a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

Table 1.A12: Representation: Mechanism exercise, estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.930	-1.015	-0.477	-0.355		
	(0.093)	(0.094)	(0.100)	(0.102)		
Female × (2001–2010)					-0.531	-0.427
					(0.168)	(0.166)
Female × (2011–2019)					-0.463	-0.337
					(0.098)	(0.100)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	350,314	350,314	350,314	350,314	350,314	350,314

The table shows the estimates from the latent model of a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist has a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.

Table 1.A13: Selection: Mechanism exercise, estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.596	-0.598	-0.123	-0.046		
	(0.077)	(0.077)	(0.080)	(0.082)		
Female × (2001–2010)					-0.524	-0.447
					(0.142)	(0.142)
Female × (2011–2019)					0.117	0.196
					(0.099)	(0.101)
Year fixed effects	no	yes	yes	yes	yes	yes
Publications & citations	no	no	yes	yes	yes	yes
Editorial positions	no	no	yes	yes	yes	yes
Fellowships & prizes	no	no	yes	yes	yes	yes
Yrs since first publication	no	no	no	yes	no	yes
Number of observations	338,600	338,600	338,600	338,600	338,600	338,600

The table shows the estimates from the latent model of a logistic regression. The data set contains author-year observations for the universe of actively publishing economists of that year. The outcome variable is an indicator equal to 1 in the years in which an economist gets a page on the English-language Wikipedia and zero otherwise. Standard errors in parentheses are clustered at the author-level.