



Who Counts as a Notable Sociologist on Wikipedia? Gender, Race, and the “Professor Test”

Julia Adams¹, Hannah Brückner², and Cambria Naslund³

Socius: Sociological Research for a Dynamic World
Volume 5: 1–14
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2378023118823946
srd.sagepub.com


Abstract

This paper documents and estimates the extent of underrepresentation of women and people of color on the pages of Wikipedia devoted to contemporary American sociologists. In contrast to the demographic diversity of the discipline, sociologists represented on Wikipedia are largely white men. The gender and racial/ethnic gaps in likelihood of representation have exhibited little change over time. Using novel data, we estimate the “risk” of having a Wikipedia page for a sample of contemporary sociologists. We show that the observed differences (in academic rank, length of career, and notability measured with both H-index and departmental reputation) between men and women sociologists and whites and nonwhites, respectively, explain only about half of the differences in the likelihood of being represented on Wikipedia. The article also enumerates both supply- and demand-side mechanisms that may account for these continuing gaps in representation.

Keywords

Wikipedia, gender gap, notability, racial/ethnic gap

Introduction

It is well known that women and people of color are underrepresented in contemporary academia and other high-status professional sectors. They are also underrepresented in discussions about and enumerations of notable persons, such as scholars, artists, or entrepreneurs, whether in reference works or news media. Wikipedia, the online encyclopedia that for many users, especially younger people, has supplanted traditional print repositories of received knowledge, has pages on the gender and racial-ethnic “gaps” on the encyclopedia itself and sponsors various projects aimed at ameliorating such gaps.¹

Previous research on the topic has focused on Wikipedia’s gender gap.² Reagle and Rhue (2011) have estimated the

proportion of female biographies on Wikipedia at 16 percent. Notable women’s odds of being omitted on Wikipedia were 2.5 of notable men’s odds, while the Encyclopedia Britannica did somewhat better with a 1.5 odds ratio. Klein and Konieczny (2015) find female representation in the online encyclopedia between 12 percent and 27 percent across human history, depending on culture and language, and actually propose to use their measure as a country-level index of gender inequality similar and complementary to measures such as the gender empowerment index. They also show that female biographies are growing exponentially on Wikipedia, with an estimated year of gender parity in 2034, although they caution that most likely the observed growth will slow down. Laouenan et al. (2018) estimate that there are at least 70,000 notable women missing from Wikipedia. Similarly, Wagner et al. (2016) have argued that Wikipedia has a glass ceiling for women. They show that Wikipedia’s women are

¹See for example https://en.wikipedia.org/wiki/Racial_bias_on_Wikipedia; https://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia; https://en.wikipedia.org/wiki/Wikipedia%3ASystemic_bias. Accessed June 4, 2018.

²Although many studies of the gender gap on Wikipedia exist, we have almost no information about racial and ethnic gaps. Two likely reasons are that race/ethnicity is difficult to identify with existing computational means and manual coding of the very big data crunched in these studies would be a monumental task.

¹Yale University, New Haven, CT, USA

²NYU-Abu Dhabi, Abu Dhabi, AD, United Arab Emirates

³Princeton University, Princeton, NJ, USA

Corresponding Author:

Hannah Brückner, NYU-Abu Dhabi, P.O. Box 129188, Abu Dhabi, AD 129188, United Arab Emirates.

Email: hb63@nyu.edu



on average more notable than its men and interpret this to mean that editors impose a higher notability threshold on women than men.³ Analyzing 893,380 biographical articles on the English language Wikipedia, they also find only 16 percent pertaining to women.

Yet what we often do not know precisely is whether notable women and minorities are rare on Wikipedia and its ilk or just rare in general. This paper aims to provide a tentative answer to this question for the academic discipline of sociology, focusing on American sociologists. Furthermore, it is usually unclear whether Wikipedia is just transmitting inequality already present in the world or introducing additional inequality by gatekeeping or other mechanisms. This is not a trivial distinction. Wikipedia's relatively open structure provides ample opportunities for mission-driven contributors, more than in traditional reference works based on expert-led entries, but arguably, Wikipedia cannot be held responsible for eliminating underrepresentation of groups and categories that simply mirror real-world inequalities. Such a stance might even conflict with the encyclopedia's policy on editor's "neutral point of view," of which more is described in the following.

The portions of Wikipedia dedicated to academics can in the first instance be conceptualized structurally, in both horizontal and vertical dimensions. The horizontal extension of Wikipedia, which changes over time and as links expand, is important to capture and make analytically tractable.⁴ Vertically, Wikipedia maps scholarship, itself a would-be map of the world. The relative congruence between two of the three conceptual levels—Wikipedia and scholarship—can be studied: We can ascertain whether the status quo of modern scholarship in given subareas is adequately represented on the online encyclopedia. It is important to note that status quo is already distorted by well-known inequalities: Despite decades of progress, for example, female and minority academics remain underrepresented in top-tier university jobs, journal boards, scholarly associations, and other positions of power, while the forms of knowledge associated with women and racial/ethnic minorities often remain underfunded and undervalued in both university departments and elite publishing outlets (Ferber 1986; Rossiter 1993). As we move to consider Wikipedia's picture of academic scholarship, however, the question arises: Are the inequalities of scholarship simply

³The article also argues that women's biographies focus more on personal and familial aspects and less on their contributions to the public sphere. In addition, the authors find linguistic bias and structural gender differences in how articles about women are linked to others.

⁴The extendable character of formations of knowledge is a reason for pessimism with respect to gender equality for in the nature of Robert Merton's (1968) Matthew effect, the masculinist structures that underpin early and existing Wikipedia entries can be expected to continue to generate further unbalanced linkages and through this mechanism alone have lasting impact.

reproduced, or are they amplified and reshaped? Does Wikipedia recreate the same patterns of underrepresentation and exclusion, or do distinctive patterns emerge?

Accordingly, we define gender and racial-ethnic minority gaps on Wikipedia as potential multipliers of inequalities produced in the extra-digital world. To approximate these multipliers, we compiled data on a population of American sociologists and matched them with what is found on Wikipedia—which enables us to estimate the risk of being represented on Wikipedia as well as differences between both male and female sociologists and whites and minorities. Existing studies of Wikipedia almost always either analyze only Wikipedia data or compare Wikipedia with other encyclopedias and reference works; we know of no other study of Wikipedia that uses the concept of an at-risk population⁵ to study Wikipedia content. We build a model of the probability of having a biography on Wikipedia that includes measures of what Wikipedia defines as notability for academics and show that women and people of color are less likely to have a Wikipedia page even after controlling for notability. This is of course a scientifically conservative measure—for just as is true of academia, Wikipedia's conceptualization of notability engages measurement strategies that do not adequately capture the actual achievements and contributions of women and minorities.

Sociology as a discipline, and American sociology in particular, is well suited as an initial site in which to systematically study gender, race, and the crowdsourced representation of academic notability. In the discipline's contemporary U.S.-based incarnation, women and minorities have significant demographic representation. In addition, recent tendencies in the field actually can be expected to have enhanced the representation of women and minority scholars in the pages of Wikipedia. There is, first of all, ongoing academic work that points out the problem in a way specific to the sociological field. The discipline as organized in the United States has also mounted organizational interventions into improving the site, including when the American Sociological Association (ASA) undertook a wide-ranging "real utopia" Wikipedia initiative under the ASA presidency of Erik Olin Wright (Wright 2011). In contrast to some academic disciplines, therefore, as discussed further in the following, sociology could be expected to do a relatively good job representing the professional achievements of women and minorities.

Notability

The measurement and production of scholarly notability and public repute is important and contested terrain in all academic disciplines, although there is differential agreement

⁵We borrow the term *at-risk population* from event history analysis to capture the idea that every academic could in principle have a Wikipedia page.

among scholars in some disciplines rather than others. The organs of academic self-governance, including recruitment committees and tenure and promotion bodies, routinely grapple with the question of how to assess and measure the quality of scholarship. It is an intrinsically difficult question. And far from being a purely academic discourse, questions of notability, quality, and accuracy of scholarship and knowledge pervade the public sphere. Many journalists reporting on academics and scientific knowledge, for example, have given up on separating the wheat from the chaff; they are also editorially guided to maintain a “balance” between opposing views and/or to tell “all sides of a story”—no matter how far out from the mainstream or distant from the real such views or “sides” may be.⁶

Wikipedia has its own version of this policy regarding such contests and assessments. Its editors are somewhat contradictorily enjoined to use “reliable sources,” *not* rely on “original research,” and maintain a “neutral point of view.” In 2001, in an influential statement that Wikipedia dubs the “original formulation,” co-founder Jimmy Wales sought to reconcile some of these approaches:

A general-purpose encyclopedia is a collection of synthesized knowledge presented from a neutral point of view. To whatever extent possible, encyclopedic writing should steer clear of taking any particular stance other than the stance of the neutral point of view. . . . Perhaps the easiest way to make your writing more encyclopedic, is to write about what people believe, rather than what is so. If this strikes you as somehow subjectivist or collectivist or imperialist, then ask me about it, because I think that you are just mistaken. What people believe is a matter of objective fact, and we can present that quite easily from the neutral point of view.⁷

The “no original research” or no primary sources rule in particular, while understandable given Wikipedia’s initial approach, conflicts with the ambition to adequately cover scholars and academic topics. This rule alone prevents many individuals and topics that have been marginalized from regaining a secure foothold on Wikipedia—or entering it at all (Luo, Adams, and Brückner 2018). It has also made for bemusement, frustration, and even heartache among people who, discovering their Wikipedia entries to be rife with slander and factually incorrect statements about them, sought to remove such material.⁸

⁶“In objective journalism, stories must be balanced in the sense of attempting to present all sides of a story” (<https://ethics.journalists.org/topics/balance-and-fairness/>).

⁷<https://web.archive.org/web/20010416035757/http://www.wikipedia.com/wiki/NeutralPointOfView>. Most recently accessed June 4, 2018.

⁸The most famous case is undoubtedly that of the novelist Philip Roth. See his ingeniously roundabout solution to correcting misstatements on his own Wikipedia page: <https://www.newyorker.com/books/page-turner/an-open-letter-to-wikipedia>. A more challenging case is that of the sociologist Frances Fox Piven.

Wikipedia’s rules—as they are currently formulated—encode several additional challenges for academics. First, the capacity of people to assess intellectual quality may be less true for highly specialized fields, where only a handful of intellectuals may be deeply familiar with the cutting edge of the science in question, or for disciplines, including a number of the social sciences, that have no universally accepted theoretical and methodological core. Assessments of quality are at least locally variable in such a space. Second, the assessment of reliability of sources is at least to some extent a matter of perspective. Although there are certainly more or less informed or expert points of view about the reliability of a given academic source, there is, of course, no neutral viewpoint independent of the observer. And in the absence of expertise among editors themselves, how to assess which among those sources is the credible expert?

In its subarea geared toward academics, Wikipedia approaches some of these linked quandaries via a specific, codified version of the notability criterion, maintaining a policy that it also refers to as “the professor test.” Here, in virtual space, is where the orientations of tenure and promotion committees and Wikipedia editors meet. The Wikipedia criteria for notability of academics include references to making a “significant impact on the field” and being an “elected member of a highly selective and prestigious scholarly society or association” or the “highest-level elected or appointed administrative post at a major academic institution,” with ensuing discussions on how to establish whether or not these criteria are met.⁹ There are gendered and racialized patterns and criteria already embedded in these judgments. Some of the highly prestigious academic societies overwhelmingly elect white men into their ranks (Ngila et al. 2017), for example, and all else being equal, women and minorities are underrepresented in research universities as holders of named chairs and even more so among university presidents and provosts.

Data on citation counts, journal impact factors, and other relatively easily numerated measures of dissemination of published work are often regarded as more even-handed indicators; they are now often included in academic

See the talk page on Piven’s Wikipedia page and discussions at https://en.wikipedia.org/wiki/User_talk:Fannielou and https://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion/Frances_Fox_Piven. Through an intermediary user, Piven asked to have the Wikipedia page about her deleted or reduced to a stub when, after having been targeted by conservative commentator and media personality Glenn Beck, she was beleaguered by abusive comments and death threats. The intermediary was banned from Wikipedia, and in an awkward compromise, some of the offensive material was ultimately removed and the page protected from edits by new and anonymous users.

⁹See [https://en.wikipedia.org/wiki/Wikipedia:Notability_\(academics\)](https://en.wikipedia.org/wiki/Wikipedia:Notability_(academics)) (accessed June 4, 2018) for information on the professor test and initial criteria for evaluating academic notability.

review dockets or contractual documents that specify performance expectations. They are routinely referred to in discussions about Wikipedia content as well in attempts to satisfy the professor test. An advantage of this approach is that bibliometric measures, for example on Google Scholar, are easy to find and increasingly accessible to nonacademics. They also seem to offer an initial and relatively uncontroversial way out of the dilemma of establishing notability. As indices of notability, however, citation counts are problematic, even biased, in a number of ways. Open-access online sources underrepresent people whose careers evolved before the advent of the Internet age. The impact of books is often not counted (and beyond that is systematically less calculable than that of articles; Clemens et al. 1995). Women and scholars of color receive fewer citations than white men in part simply because of who they are perceived to be and in part because they may be less likely to cite themselves. Other factors held constant, they are published less in top peer-reviewed journals (e.g., Chakravarthy et al. 2018; Hartman et al. 2013; Lariviere et al. 2013; Long 1992; Ray 2018; Xie and Shauman 1998).¹⁰ Women and minorities are also concentrated in disciplines with overall lower citation counts (Merritt 2000). Furthermore, citation has also been shown to be in healthy part reputational and professionally strategic, which replicates and deepens discriminatory processes (Bornmann and Daniel 2008).

Far from helping Wikipedia contributors with the improvement of knowledge, therefore, Wikipedia's treatment of academic notability—the so-called professor test—is among the factors that transmit existing inequalities into the encyclopedia and magnify them informationally, taking us further away from an approach that would be both more accurate and increasingly inclusive.¹¹ Nonetheless, we adopt the approach of studying "notability" empirically as an important first step in understanding, in the most rigorous way, how crowdsourced assessments of academic merit do or do not recognize women and people of color. In other words, we pose the following question: Are Wikipedia's criteria for academic notability actually applied by and in Wikipedia itself?

¹⁰Some of the research of gender disparities in productivity and impact shows that these differences have declined over time and among younger cohorts of scholars (e.g., Van den Besselaar and Sandström 2016; Xie and Shauman 1998) with the declining gender gap in the academy. However, as we show in the following, gender differences in citations counts are still substantial in a population of academics in R1 universities.

¹¹Laouenan et al. (2018) use Wikipedia to construct a multilingual database of notable individuals spanning millennia. While their approach to notability differs from ours, they also find women to be systematically underrepresented.

Methodology and Data

Research on Wikipedia and other online resources is generally limited to exploring the contents of these resources.¹² We construct, in contrast, a population "at risk" of representation on Wikipedia—namely, a real-world sample of faculty members in sociology—and compare this population with who is actually represented on Wikipedia. As part of a larger project, we compiled faculty data from the top sociology departments in the United States using all sociology departments located in Research 1 universities ($N = 96$).¹³ The discipline as a whole is large enough to be difficult to access with sufficient detail. While labor-intensive, however, it was possible to scrape and manually code data from a smaller number of sociology departments; the idea behind using the highest ranked departments was that scholars in these departments would be more likely to be at risk to have a Wikipedia page, which, we surmised, is rare among academics, hence focusing on contexts in which a research-active academics cluster will provide more analytical information. Clearly, however, such a sample is not meant to be representative of the discipline as a whole.

This resulted in data on 2,978 faculty working in these departments in the fall of 2014 (hereafter referred to as R1 sample). It should be noted that this strategy excludes sociologists working in liberal arts colleges and other institutions that do not have graduate programs and are not ranked as research universities (see the following for more discussion). Gender, race/ethnicity, rank, and year of PhD were hand-coded using all available information on the faculty members' webpages and elsewhere on the web. We also collected keywords related to the research program of the academics where possible. Rank (as noted on the department's webpages) distinguishes among named or distinguished professors, emeriti, full professors, associate/assistant professors, and a residual category for adjuncts, lecturers, and so on. The sample is 43 percent female (somewhat less than in the discipline at large, which is to be expected as women are less likely to hold jobs in highly ranked universities¹⁴). We approximated race/ethnicity using photographs, names,

¹²There are some exceptions. Elvebakk (2008) compares philosophers on Wikipedia with two other online databases that were compiled by academics. Lam et al. (2011) compare movies with entries on Wikipedia with another database, and Halavais and Lackaff (2008) compare Wikipedia against the distribution of topics of published books and two academic encyclopedias.

¹³Research 1 universities are those with doctoral programs and the highest research activity as measured by the Carnegie classification of higher education institutions (<http://carnegieclassifications.iu.edu/>). We also used various graduate program rankings and other measures of departmental excellence and chose to be inclusive with respect to these measures, meaning that if a department appeared in the top 100 of any ranking metric, we included it.

¹⁴ASA membership was 53 percent female in 2016 and has been gender-integrated for two decades (<http://www.asanet.org/research-and-publications/research-sociology/trends/asa-membership-gender>; last accessed July 31, 2018).

and curriculum vitae information, distinguishing among the categories of white, Hispanic, African American, and Asian. While we are aware that our strategy to measure race/ethnicity is not without problems of reification and misclassification, we believe that it is preferable to simply ignoring race and ethnicity, as most of the literature on Wikipedia has done.

Even this relatively large sample does not yield many academics from underrepresented minorities¹⁵: The sample is 81 percent white, 6 percent black, 5 percent Asian, 4 percent Hispanic, 1 percent other, and the remainder race/ethnicity unknown.¹⁶ Wikipedia presence is particularly low for the nonwhite groups, as discussed in the following, and we were therefore not able to estimate group differences beyond a basic white/nonwhite dichotomy. This is in part due to the sheer level of underrepresentation of minority professors in the academy, accentuated at research-intensive institutions (Etzkowitz, Kemelgor, and Uzzi 2000; Myers 2016).

Sociologists on Wikipedia

In a second step, using Wikipedia's API, we compiled data on everyone listed under the category "American Sociologist" on Wikipedia and also conducted a name search using the set of names from our sample of R1 sociologists. We matched the names from both data sets to generate a data set that allows us to estimate the probability of having a Wikipedia biography. Of the 710 sociologists found on Wikipedia in August 2014, 452 were alive at that point. Slightly more than half of these (53 percent) were also in our R1 sample. We categorized the remainder according to the reason we could not match the case. About half of these cases are in schools that are not represented in our R1 sample and are working in a heterogeneous mix of selective and nonselective colleges and state campuses. Schools that contribute more than one case to this group are Boston College with four cases and Brigham Young University with a whopping eight cases. About 10 percent are sociologists working in R1 universities who are however listed under other departments or professional schools, with no indication of an affiliation with the requisite sociology department. Another 16 percent have left academia or are working in universities abroad. A small number (5 percent) were missing from their departments'

¹⁵The proportion of PhDs awarded to nonwhite graduates has been around one-third since 1990, indicating that the R1 sample is less diverse than the profession at large. ASA membership was 73 percent white in 2015 (see <http://www.asanet.org/research-publications/research-sociology/trends-sociology/race-and-ethnicity>; accessed September 1, 2018).

¹⁶Almost none of the 76 scholars of unknown race have Wikipedia pages, and for most of them, we are also missing other data. They are therefore omitted from results pertaining to racial differences reported in the following.

webpages and were reclassified as a match. Finally, almost a fourth of the group were not sociologists according to our definition, and in many instances, there was no indication that these individuals had any connection to sociology or academia in general. This heterogeneous group includes social workers, social activists, journalists, pop psychologists, religious leaders, and a variety of academics working in other disciplines.¹⁷ Nevertheless, there is considerable overlap between our definition of a sociologist and what is found on Wikipedia, making our strategy to sample an at-risk population viable. We updated our Wikipedia sample in October 2016 and found 553 living sociologists, of which 362 (65 percent) were also in our R1 sample. Sociologists with Wikipedia pages are predominantly male (76 percent) and white (88 percent). Clearly, even compared to the already select R1 sample, Wikipedia favors white male scholars.

Figures 1 and 2 show how the demographic composition of sociologists on Wikipedia developed over time, from Wikipedia's beginnings to 2016. Figure 1 traces the cumulative frequencies in absolute numbers for white and nonwhite men and women. Figure 2 shows the proportion in each group over time. In 2002, our data contain only five sociologists—with Sherry Turkle and Manuel Castells as the nonwhite male members. Two more women (Arlie Hochschild and Pepper Schwartz) were added in 2004, and a number of nonwhite and female sociologists were added in 2005. After these fluctuations during Wikipedia's beginnings, which could be attributed to small numbers overall, the next decade shows much growth in absolute numbers but more stability than change in terms of proportions. The proportion of nonwhite males and females remained virtually constant, and the proportional increase for white women is gradual, from 17 percent in 2005 to 22 percent in 2016. It is perhaps noteworthy that Eric Olin Wright's 2011 ASA presidential Wikipedia initiative (see Wright 2011) seems to have made little impact on the numbers shown in Figures 1 and 2.

Notability Measurement

As noted previously, Wikipedia's professor test mentions citations as a sign of notability and discusses measures developed in the academy that attempt to reduce this record to a single numerical measure, including the H-index.¹⁸ In fact, as we will show in the following, the H-index and citations are frequently mentioned in notability discussions on Wikipedia. The accuracy of such measures, however, depends on building accurate publication citation records. Research on citations commonly uses databases such as

¹⁷For a critical appraisal of the representation of the discipline of sociology on Wikipedia, see Adams and Brückner (2015).

¹⁸<https://en.wikipedia.org/wiki/H-index>. The H-index attempts to measure both quantity of publications and impact. An H-index of 10, for example, indicates an author has 10 publications cited 10 or more times.

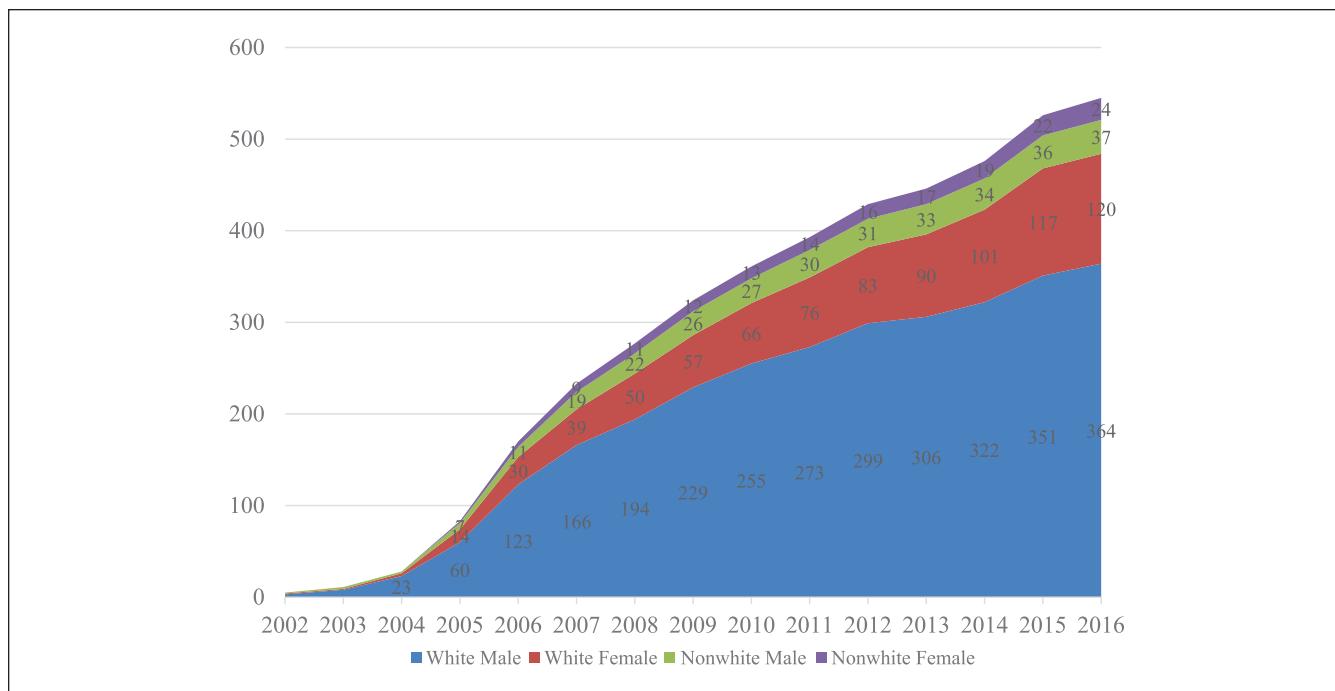


Figure 1. Cumulative frequency of American sociologists on Wikipedia, 2002–2016, by race and gender.

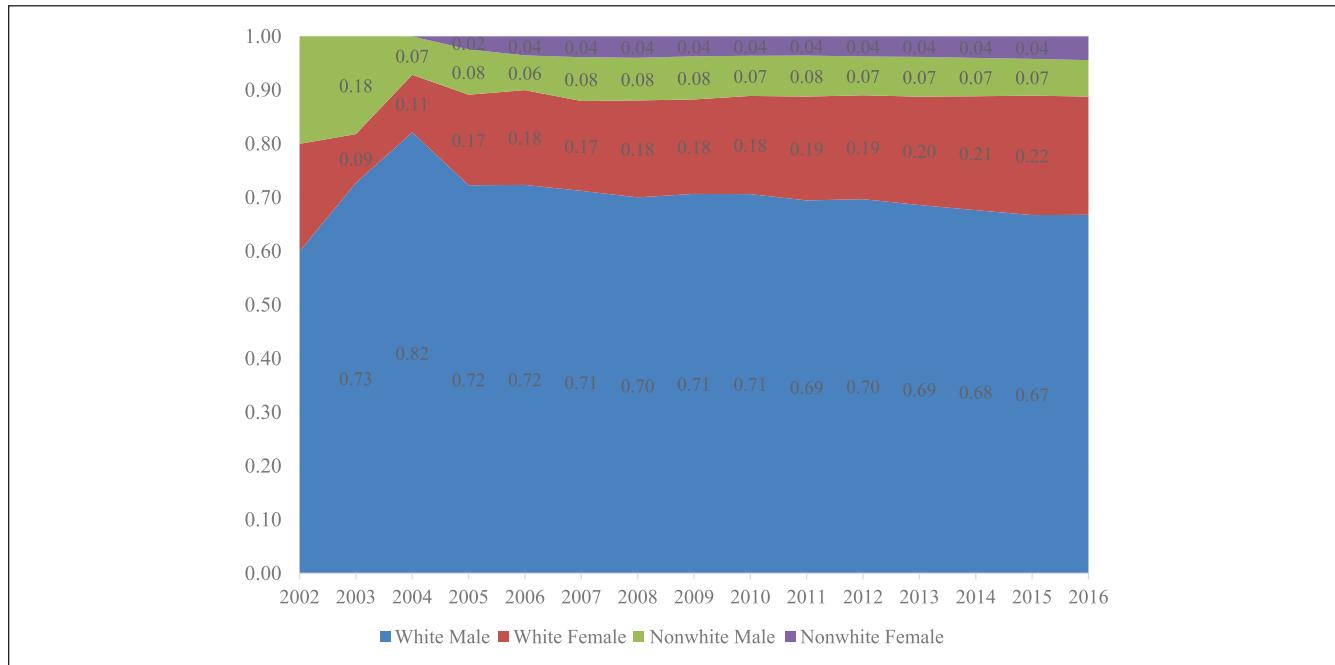


Figure 2. Composition of Wikipedia American sociologists, 2002–2016, by race and gender (proportion).

Scopus and Web of Science for convenience and accessibility, although it is known that they are biased toward articles in specific journals and do not reference materials published in other key formats, such as books. Particularly for disciplines in which books are an important publication platform, data drawn from these databases will be less accurate in

measuring impact. In addition, the databases may not be accessible for Wikipedia editors outside the academy, who would typically resort to Google Scholar, which also provides information on citations. Finally, many other databases exclude references in the nonscholarly media, which may be of interest to Wikipedia editors. We therefore decided to

Table 1. Descriptive Statistics for R1 Sociologists (N = 2,713).

	Male	Female	Minority	White	Total
Has WikiPage (%)	16	7	7	14	12
Rank (%)					
Named/distinguished	13	7	7	12	10
Emeritus	19	7	20	11	14
Full	33	26	24	33	30
Associate	21	34	26	27	26
Assistant	8	17	13	12	12
Other	5	8	9	6	7
Year of PhD (%)					
1945-1979	42	17	29	33	31
1980s	16	16	14	17	16
1990s	16	21	18	18	18
2000+	24	45	36	32	33
Top 20 department	29	24	19	31	27
H-index (mean)	20.5	14.4	14.4	19.5	17.9
H-index (median)	17	12	11	16	14
H-index (mean, logged)	2.68	2.34	2.28	2.65	2.54
Percentage of R1 sample	58	42	30	70	100

derive bibliometric data from Google Scholar to measure notability (the viability of this approach is discussed, among others, by Martin-Martin et al. 2017).

We used an automated program to run a search query for each scholar's name and gather the publication titles and their citation counts from each page of search results. Because names are not unique identifiers, the results contain not only papers by the author we were concerned with but also those by anyone who happened to share that name. Thus, our problem became how to identify papers written by the author of interest and to remove those that are not. This problem is sometimes referred to as *author-name disambiguation*. After experimentation, we developed a machine-learning approach described in the Methodological Appendix. The resulting measures correlated highly with those derived from Google Scholar profiles.¹⁹ For the H-index derived from these data, we find a correlation of .94 (.97 after some additional corrections, see Appendix), and we present results primarily based on that measure. Total number of publications and total number of citations as well as average citations per publication yield similar results, although the former two measures are somewhat noisier.²⁰ In addition, we used the *US News* ranking of graduate departments as an additional measure of notability, distinguishing the top 20 departments from others.

We use these data to test (1) whether women and minority faculty are underrepresented on Wikipedia relative to their representation in the discipline and (2) whether differences, if

any, can be explained by gender and race/ethnicity differences in notability factors alone or instead at least partially reflect inequalities endemic to the production of knowledge on Wikipedia itself. To test this latter hypothesis, we will present a logistic regression of the probability of being represented on Wikipedia as a function of academic rank, publication and citation figures, time since PhD, race, and gender.

Results

Table 1 shows descriptive statistics for the R1 sample. Male sociologists working in R1 universities are more than twice as likely to have a page on Wikipedia (16 percent) than their female colleagues (7 percent, $\chi^2 = 49, p < .001$). Similarly, white sociologists are twice as likely (14 percent) to have a page than others (7 percent, $\chi^2 = 43, p < .001$). As expected, female and nonwhite sociologists are overrepresented in lower ranks, more likely to be recent PhDs, and less likely to work in top 20 departments. Women's H-indices are significantly lower than men's, and white authors have higher H-indices than others. Differences between men and women and white and nonwhite sociologists reported in Table 1 are all statistically significant. There is a relatively small difference between minority men (13 percent) and women (10 percent) on the dependent variable, which is not statistically significant. In contrast, the large difference between white women (11 percent) and white men (24 percent) is highly significant and similar to the large differences between white men and minority scholars. As noted previously, minority scholars are rare in our data, and with larger numbers, intersectional contrasts might be statistically significant; however, these data clearly show that the inequality between white men and everybody else dwarfs other group comparisons.

¹⁹Google Scholar profiles are initiated by authors and therefore presumably cleaner than straight search results because authors can delete publications not authored by themselves and ensure completeness of the list of publications.

²⁰Results available on request from the authors.

Table 2. H-Index (Deciles) and Proportion with a Wikipedia Page, by Gender/Race.

H-index decile	White Men				Others				P
	N	Proportion	95% CI	N	Proportion	95% CI			
1	124	.06	.03 .12	240	.03	.01 .06			.17
2	120	.12	.07 .19	221	.04	.02 .08			.01
3	115	.14	.08 .22	156	.07	.04 .12			.06
4	133	.17	.11 .25	198	.08	.05 .13			.01
5	103	.16	.09 .24	132	.13	.08 .20			.56
6	182	.14	.10 .20	198	.10	.06 .15			.16
7	126	.21	.15 .30	105	.10	.05 .17			.01
8	178	.26	.20 .33	129	.17	.11 .25			.07
9	194	.32	.25 .39	123	.31	.23 .40			.84
10	233	.54	.47 .61	70	.43	.31 .55			.10

Table 2 shows that the probability of having a Wikipedia page increases with the H-index (shown in deciles). We compare white men to all other groups in Table 2, reporting proportion, binomial 95 percent confidence intervals for the proportion, and number of observations. The last column reports statistical significance of the difference between white men and others. In the bottom decile of the H-index, white men's likelihood to be on Wikipedia is 6 percent compared to 3 percent for others; in the top decile, the likelihood is 54 percent for men and 43 percent for others. Across the distribution of the H-index, white men are more likely to have a Wikipedia page, often significantly so, indicating that notability as measured here does not fully account for group differences.

Some of the literature on the gender gap on Wikipedia shows that women with Wikipedia pages are more notable than men with Wikipedia pages (e.g., Wagner et al. 2016), but this does not seem to hold for sociologists. Examining only sociologists with Wikipedia pages, men's median H-index (27) is higher than women's (22), as is the 25th percentile with 16 versus 13 and the 75th percentile with 43 versus 36. Similar gaps obtain for the differences between whites and nonwhites.

Table 3 reports results from a logistic regression on the probability of having a Wikipedia page.²¹ Here we use a logged version of the H-index to adjust for a strong right skew in the H-index distribution. Women's estimated odds of having a Wikipedia page after taking into account differences in rank, length of career, and notability measured with H-index and departmental reputation are still 25 points lower than men's. Similarly, the odds of nonwhite sociologists are 28 points lower than their white colleagues. In short, the observed differences between men and women and whites and nonwhites explain only about half of the differences in the likelihood of being represented on Wikipedia.

²¹Results are based on matching with what was on Wikipedia in October 2016. Our first Wikipedia extract from October 2014 yields similar results (available per request).

Table 3. Logistic Regression.

	Odds Ratio	p	95% CI	
Nonwhite	.72	.042	.53	.99
Female	.75	.039	.56	.99
Logged H-index	3.74	.000	2.91	4.82
Rank				
Full professor (base)	1.00			
Named/distinguished	2.07	.000	1.50	2.85
Emeritus	1.20	.347	.82	1.75
Associate professor	.48	.011	.28	.84
Assistant professor	.10	.028	.01	.78
Other	1.39	.482	.55	3.51
Year of PhD				
Pre 1980 (base)	1.00			
1980s	.91	.605	.65	1.29
1990s	.82	.324	.56	1.21
2000+	.86	.632	.47	1.58
Top 20 department	1.24	.106	.96	1.61

Note: Odds ratios that are significantly different from 1 are bolded. N = 2,913. LR = 509. Pseudo R² = .23.

We tested for interaction effects between gender/race and the H-index measure as well as rank and year since PhD since it is often the case that women and minorities have to be "better" than their male/white counterparts to achieve recognition. None of the interaction effects was significant,²² although it should be noted that we are working with relatively small numbers given how rare it is for members of these subpopulations to have a Wikipedia page.

Page Deletion Analysis

It is important to note that the missing women and minority scholars could be missing for two reasons: Either they might have been added at some point and then deleted, or they

²²Results available on request.

never commanded the attention of Wikipedia contributors in the first place. Over the course of our work on the project, we found and received credible anecdotal evidence that editors intent on adding notable women and minority persons were harassed and even banned from Wikipedia and that pages contributed by these editors had been deleted. On the basis of this evidence, we decided to explore the question of whether broader dynamics were at issue. We therefore added an analysis based on an archive of page deletion discussions related to academics.²³ Page deletions on Wikipedia are usually processed as follows: An editor proposes to delete a specific page, and that proposal is open for discussion for some amount of time. Wikipedia administrators²⁴ then make a decision based on the discussion. In some cases, a “speedy deletion” is undertaken if the deletion is deemed to be uncontroversial.

We retrieved about 90,000 comments related to 6,323 deletion discussions threads and coded gender and the result of the discussion. It is important to note that these discussions are related to academics in general. In fact, no details about the person, such as discipline or occupation, are preserved when the page is deleted, except for the name. Gender was coded using an API²⁵ with additional manual coding for about 1,000 names the API could not classify. We were able to identify gender of the person whose page was flagged for deletion for 97 percent of the threads. We were not able to identify race/ethnicity with the available information, and so this portion of our analysis focuses on gender only.

More than half of the discussions threads resulted in deletion (56 percent, which includes 87 cases where the outcome was merging with another page or redirection). We coded the outcomes keep, no consensus, and withdraw (which refers to the original deletion proposal being withdrawn) as keep. Only 17 percent of the threads are about women academics. However, pages about women were not more likely to be deleted than pages about men (45 percent and 44 percent, respectively, $\chi^2 = .82, p = .365$). Note that we cannot control for rank or notability here and that it is possible women academics’ pages are more likely to be deleted than those of comparably notable men. Nevertheless, these particular data suggest that the main story is that women are less likely to appear in the first place.

It is noteworthy that almost exactly half of the discussion threads contained references to the H-index or to citations, indicating that these measures are quite frequently mobilized as argument for or against deleting a page. Hence, the deletion discussion data suggest the validity of our strategy to

²³Retrieved from Wikipedia March 1, 2016. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Deletion_sorting/Academics_and_educators/archive.

²⁴Wikipedia administrators are Wikipedia editors selected through a “community review process” described in detail in Jemielniak (2014; see also <https://en.wikipedia.org/wiki/Wikipedia:Administrators>).

²⁵<https://www.behindthename.com/api/>.

measure notability as seen through the eyes of Wikipedia editors.

Discussion and Conclusions

The empirical work presented in this paper has primarily provided documentation and estimates of the extent of underrepresentation of women and people of color on academic Wikipedia. We show that the observed differences (in academic rank, length of career, and notability measured with both H-index and departmental reputation) between men and women sociologists and whites and nonwhites explain only about half of the differences in the likelihood of being represented on Wikipedia. However, the descriptive work on the presence and degree of underrepresentation does not speak much to the mechanisms that may account for it. In this final section, we discuss some of these potential mechanisms to situate the analysis and clarify further research questions to which it gives rise.

It should first be granted that scholarly knowledge, while historically produced in an environment that is far from gender- or race-neutral, has no direct, unmediated relationship to individual or even aggregated preferences. At least according to the tenets of the Enlightenment, the relevance of disciplinary knowledge is not structured in this sense. The insights drawn from English literature and psychology, fields today practiced mostly by women, and those drawn from physics and economics, fields practiced predominantly by men, should be equally relevant to all individuals. The work conducted within the newer disciplines of Gender and Sexuality Studies, for example, or African American Studies, is putatively pertinent to academics of any gender and sexual orientation, race, or ethnicity. Therefore, even if Wikipedia contributors are mostly white men, as they currently are (Wikimedia Foundation 2011), underrepresentation of women scholars and scholars of color in the pages of Wikipedia itself is not a necessary or logical consequence. Nonetheless, it obtains in the representation of the discipline of sociology, as we have shown, and sociology is a discipline whose demographic characteristics and recent history of specific political and institutional commitments would have been expected to generate an alternative and more inclusive outcome.

One cluster of potential mechanisms emanates from the supply side, as it were. Women and people of color may be less likely than white men to act on behalf of self in this arena. There are likely to be, in other words, systematic gender and racial/ethnic differences in the desire and capacity to promote oneself online.²⁶ It has been argued in scholarly and popular literature, for example, that women in particular are more uncomfortable with self-promotion (Gilligan 1982;

²⁶One must be careful with such arguments, however, since such self-promotion would be viewed as a conflict of interest in the rules governing Wikipedia as it would preclude a “neutral point of view” (https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)

Sandberg 2013). In addition, backlash against women's perceived self-promotion has frequently been documented (e.g., Kanter 1977; Rudman and Phelan 2008), and that backlash is in part responsible for women's reluctance to self-promote (Moss-Racusin and Rudman 2010).

It is also possible that women scholars and scholars of color are simply too busy to devote time to organizing their own virtual representation, in part because of the known "cultural taxation" of women and minorities in the academy (Joseph and Hirschfield 2011). However, data from a very large survey of Wikipedia readers and editors indicate that women were equally or less likely than men to give lack of time as a reason for not contributing (any longer) to Wikipedia. Rather, women were considerably more likely than men to agree with items measuring actual or potential conflict with others as reasons for not contributing to Wikipedia (Collier and Bear 2012). Women and people of color in academia might also be less likely to have a devoted following of individuals who are able to mobilize to create a biographical page for them. Finally, and relatedly, women and scholars of color might be disproportionately excluded from academic networks (see Etzkowitz et al. 2000) that include those contributing to Wikipedia. These are all empirically researchable questions.

A second set of major mechanisms focuses on the demand side of the equation, suggesting that Wikipedia gatekeepers are likely to apply to contributors the same gatekeeping dynamics that amplify race and gender inequality in other social contexts, virtual or otherwise. Dismissiveness and hostility toward women and minorities in virtual space does occur and has frequently made news in recent years (Bartlett et al. 2014; Buni and Chemaly, 2014; Eckert and Steiner 2013; Tripodi 2017). Lam et al. (2011) found that articles with a high number of female editors were more likely to be controversial, female newcomers to Wikipedia were more likely to see their edits reversed, and female editors were more likely to be indefinitely banned from Wikipedia. Conflicts on Wikipedia over content have resulted in some long-time female contributors being banned for life, while their male opponents were slapped on the wrist in spite of uncivil or threatening behavior (Auerbach 2014).

Such conflicts have often been related not to debate in some neutral sense but to specifically feminist contributions, whether actually feminist in any way or merely labeled as such. And because they are well publicized, they reinforce the assumption that Wikipedia is not a broadly hospitable digital space. Sociology as a discipline might be a particularly fraught arena in this perceptual context because research on inequalities of gender and race is both prominent in the

Wikipedia:Conflict_of_interest). Suspected conflict of interest will result in page deletion should a Wikipedia page be challenged. Editors may go to great lengths to uncover self-promotion, including looking up the relevant university's IP address and flagging as suspicious edits coming from the person's home institution.

discipline and disproportionately engaged in by women and minority sociologists. This forms part of another demand-side explanation, which basically holds that just as in the tech sector overall, the young men who largely exercise editorial control over Wikipedia content are at best uninterested in and often hostile to representing the contribution to knowledge by scholars who are overlooked based on their position or prevailing practices in academia. We do not have the quantitative data to assess this agent-based hypothesis, but it is potentially testable.

These supply- and demand-side mechanisms—by no means mutually exclusive—bear further investigation. They are related to contending explanations of women and minorities' underrepresentation in higher management positions at work and in the public sphere and the long-standing arguments surrounding them (e.g., Jellison 1987). Yet they also differ insofar as they include characteristics of the digital sphere, such as algorithmic bias (Baeza-Yates 2018) or emergent forms of action deriving from online anonymity. Nor should they be expected to work in completely parallel fashion for gender and race/ethnicity. In future, given enough cases to do the relevant quantitative work, it would be interesting to examine specific Wikipedia pages on which somewhat different supply- and demand-side mechanisms causing the gender and race/ethnicity gaps were operating.

In the era of #MeToo and #BlackLivesMatter, the underrepresentation of women and people of color in the media and as recognized voices in the U.S. public sphere has given rise to compensatory initiatives, such as the *New York Times*'s new feature of obituaries of notable but historically "overlooked" nonwhite, nonmale people.²⁷ Wikipedia editors, following the platform's "no original research" policy, might be hard pressed to find bibliographical material on people who have been systematically excluded from public representation—for at least some part of the multiplication of gender and racial bias is rooted elsewhere in the public domain, in the "reliable sources" writing about "what people think," which constitutes raw material for Wikipedia entries. Although academics, after all, are notable because of their contributions to knowledge, and that, in most cases, is available through available published work, Wikipedia policy continues to preclude synthesis and analysis of an academic's work itself to fill in the blanks in the public and scholarly spheres because such analysis is viewed as a violation of the primary source/no original research policy. At present, Wikipedia editors can only use secondary sources to write about someone's work.²⁸ While this policy does not bode

²⁷<https://www.nytimes.com/interactive/2018/obituaries/overlooked.html>.

²⁸The relevant rule is this: "Do not analyze, evaluate, interpret, or synthesize material found in a primary source yourself; instead, refer to reliable secondary sources that do so" (https://en.wikipedia.org/wiki/Wikipedia:No_original_research; last accessed July 31, 2018).

well for Wikipedia's potential as the sort of "real utopia" that Erik Olin Wright (2011) envisioned, it should be noted that discerning Wikipedia editors could go out of their way to unearth reliable sources on the missing women and people of color—though perhaps not as easily as launching a Google search—for historians and other scholars have long compiled materials that document the contribution of the overlooked to all academic disciplines. Such efforts, however systematic, may trigger disingenuous or genuine but misbegotten efforts by the largely white male Wikipedia editorial group or network to invoke the "neutral point of view" policy, discrediting efforts to document such contributions on Wikipedia. More robust participation by a demographic and intellectual cross-section of academics in the representation of academic disciplines on virtual platforms is one possible antidote.

Methodological Appendix

Method for Name Author-name Disambiguation

We began by trying an approach presented by Ruths and Zamal (2010), who propose applying a number of filters to the search results that remove publications unlikely to represent the work of the individual in question based on other information available about the author. The simplest of these removes all publications occurring too far in advance of the year of the author's PhD. Other filters discard publications whose author names are inconsistent with that of the author in question or in conflict with each other (JS Doe and JM Doe are unlikely the same individual). While in theory this should work, in practice we found that many scholars are inconsistent with the names under which they publish and parsing errors on the part of Google when compiling results on the web can also lead to inconsistencies in author name on otherwise valid publications. The most effective filter used by Ruths and Zamal was based on vocabulary extracted from an individual's webpage or CV. This filter kept only those publications containing a significant number of terms from this vocabulary set. For the individuals in our data set, we found that this type of information was not always available as many individuals only had abbreviated or outdated CVs available online or none at all.²⁹ In addition, the filters were too restrictive and led to measures that significantly underestimated citations.

After examination of the citation data using filters and comparison with curated Google Scholar profiles, we decided to go another route: to collect all items resulting from an author search using only the name and applying a disambiguation method afterward. To achieve this, we used a large set of titles for which we knew the domain and used it to build a classification algorithm. Google Scholar Metrics

keeps a list of the top publications in eight different subjects, including Social Science, and the top-cited articles in the past five years (specifically those that contribute to the publication's H-5 index) for each. From this we scraped 196,916 titles. Of these, 23,702 were from journals classified under Social Science or one of its subcategories. An additional 40,103 titles³⁰ were collected from the curated Google Scholar profiles of 720 sociologists. From all of the titles a set was constructed containing titles from the social science journals and Google Scholar profiles, balanced by an equal number of titles randomly selected from the other domains. This was then split into a training set and a testing set. The titles were then transformed into a matrix of binary features for the occurrence of tokens (words) in the titles. English stopwords were removed. As domains can overlap substantially, it is much easier to classify a title as belonging or not belonging to a single domain rather than selecting a particular domain from a larger set. Thus, the dependent variable for our classification model was a binary indicator of membership within the domain of social science/sociology. After testing a number of classification algorithms, we selected the one that provided the greatest accuracy in predicting a title's membership while maintaining specificity and sensitivity. We chose a Bernoulli naive Bayes model that uses the binary occurrence of tokens in the text as predictors, rather than frequency, and thus performs well in classifying short texts. When run on the test set, our model performed well in overall accuracy (86 percent), precision (84 percent), and recall (89 percent).

The following table shows the correlations of indices between the curated Google Scholar profiles and publication lists created with various filtering methods. For all indices, the publication record filtered using the Bernoulli naive Bayes classifier and PhD year filter provides the closest results to the curated set.

Filters Applied	H-Index	Average Citations	Total Citations	Number of Publications
PhD year filter only	.7557	.8869	.6480	.5099
PhD year and name filters	.7169	.8115	.7596	.6207
PhD year, name, and keyword filters	.8458	.8302	.9246	.7545
Bernoulli naive Bayes classifier and PhD year filter	.9446	.8870	.9552	.7908

Our method improves on that proposed by Ruths and Zamal by eliminating the name-based filters, which exclude

²⁹We were not able to collect keyword data on about 700 scholars.

³⁰Originally 62,233 titles, but those without complete citation information were removed as many profiles have messy tails with ill-parsed entries whose inclusion would not improve our classifier.

valid publications in cases where author names are parsed incorrectly by Google and where an author may cite under inconsistent names over the course of their careers. This filtering method might also introduce a gender bias as it is more common for women to add a name or hyphen when they marry. Additionally, it improves on the vocabulary filter by applying a single and standard filter for each author within a domain, thus eliminating the need for collection of CVs, which can be difficult to track down, as well as the errors in measurements introduced when these are incomplete or missing. Furthermore, our title classifier provides a more sophisticated and accurate prediction than shared vocabulary counts. For example, in the title “RNA Molecular Weight Determinations by Gel Electrophoresis under Denaturing Conditions, a Critical Reexamination” (Lehrach et al. 1977), the tokens *determinations*, *conditions*, and *critical reexamination* are all commonly used in sociology and might be likely to appear in any given sociologist’s CV, but in our model, the presence of the additional tokens *RNA molecular weight* and *gel electrophoresis* would tip the scales appropriately to reject the title.

The use of this method relies on the assumption that two sociologists (or any academics within a given discipline) will not share a name. This assumption is bound to fail in some cases, but for known exceptions, the publication records for these individuals can be corrected manually. It might be possible to solve this problem using a clustering algorithm to tease apart separate bodies of work by authors sharing a name, but with only titles and no a priori knowledge about the number of clusters (i.e., the number of distinct authors sharing a name) or the relative cluster sizes, we were not able to achieve this. With access to full documents (and thus a much larger set of features) for each search result, cluster analysis could be an area for future research in improving solutions to the author-name disambiguation problem. To gauge the seriousness of the problem, we correlated the derived bibliometric with a variable derived from 2010 census data that measures the frequency of the scholar’s last name in the population.³¹ This was available only for the 1,000 most common names, and we set the frequency for other names to zero. A significant correlation indicates that the measures are affected by incomplete name disambiguation. The correlation was indeed significant, albeit only for men.

To address this problem, we ran a model that regressed the H-index as derived previously on scholar’s characteristics (including rank and year since PhD) and examined the outliers from this model. We also checked data for scholars with common names and those with names that occurred more than once in our R1 sample. For these cases, we manually corrected the citation data, using Google Scholar profiles where available. The resulting corrected data showed

³¹https://www.census.gov/topics/population/genealogy/data/2010_surnames.html.

no correlation with the name frequency as measured by the census, and the correlation of the resulting H-index correlates with the Google Scholar profile data with .97.

Acknowledgments

We are grateful to our research assistants, Jesse Einhorn and Sima Shabaneh, and to Peter Bearman for helpful comments on earlier versions of the paper.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge support from NSF Grant No. 1322971.

References

- Adams, Julia, and Hannah Brückner. 2015. “Wikipedia, Sociology, and the Promise and Pitfalls of Big Data.” *Big Data & Society* 2(2):1–5.
- Auerbach, David. 2014. “Encyclopedia Frown.” *Slate.com*. Retrieved August 29, 2018 (http://www.slate.com/articles/technology/bitwise/2014/12/wikipedia_editing_disputes_the_crowdsourced_encyclopedia_has_become_a_rancorous.html).
- Baeza-Yates, Ricardo. 2018. “Bias on the Web.” *Communications of the ACM* 61(6):54–61.
- Bartlett, Jamie, Richard Norrie, Sofia Patel, Rebekka Rumpel, and Simon Wibberley. 2014. “Misogyny on Twitter.” Demos. Retrieved August 29, 2018 (https://www.demos.co.uk/files/MISOGYNY_ON_TWITTER.pdf).
- Bornmann, Lutz, and Hans-Dieter Daniel. 2008. “What Do Citation Counts Measure? A Review of Studies on Citing Behavior.” *Journal of Documentation* 64(1):45–80.
- Buni, Catherine, and Soraya Chemaly. 2014. “The Unsafety Net: How Social Media Turned against Women.” *The Atlantic*. Retrieved November 18, 2018 (<http://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/>).
- Chakravarthy, Paula, Rachel Kuo, Victoria Grubbs, and Charlton McIlwain. 2018. “#CommunicationSoWhite.” *Journal of Communication* 68(2):254–66.
- Clemens, Elisabeth S., Walter W. Powell, Kris McIlwaine, and Dina Okamoto. 1995. “Careers in Print: Books, Journals, and Scholarly Reputations.” *The American Journal of Sociology* 101(2):433–94.
- Collier, Benjamin, and Julia Bear. 2012. “Conflict, Criticism, or Confidence: An Empirical Examination of the Gender Gap in Wikipedia Contributions.” Pp. 383–92 in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. New York: ACM.
- Eckert, Stine, and Linda Steiner. 2013. “(Re)triggering Backlash: Responses to News about Wikipedia’s Gender Gap.” *Journal of Communication Inquiry* 37(4):284–303.

- Elvebakk, Beate. 2008. "Philosophy Democratized?" *First Monday* 13(2).
- Etzkowitz, Henry, Carol Kemelgor, and Brian Uzzi. 2000. *Athena Unbound: The Advancement of Women in Science and Technology*. Cambridge: Cambridge University Press.
- Ferber, Marianne A. 1986. "Citations: Are They an Objective Measure of Scholarly Merit?" *Signs: Journal of Women in Culture and Society* 11(2):381–89.
- Gilligan, Carol. 1982. *In A Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.
- Halavais, Alexander, and Derek Lackaff. 2008. "An Analysis of Topical Coverage of Wikipedia." *Journal of Computer-Mediated Communication* 13(2):429–40.
- Hartmann, William E., Eric S. Kim, Jackie H. J. Kim, Teresa U. Nguyen, Dennis C. Wendt, Donna K. Nagata, and Joseph P. Gone. 2013. "In Search of Cultural Diversity, Revisited: Recent Publication Trends in Cross-cultural and Ethnic Minority Psychology." *Review of General Psychology* 17:243–54.
- Jellison, Katherine. 1987. "History in the Courtroom: The Sears Case in Perspective." *The Public Historian* 9(4):9–19.
- Jemielniak, Dariusz. 2014. *Common Knowledge? An Ethnography of Wikipedia*. Stanford, CA: Stanford University Press.
- Joseph, Tiffany D., and Laura Hirschfield. 2011. "Why Don't You Get Somebody New to Do It?" Race and Cultural Taxation in the Academy." *Ethnic and Racial Studies* 34(1):121–41.
- Kanter, Rosabeth M. 1977. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *The American Journal of Sociology* 82:965–90.
- Klein, Maximilian, and Piotr Konieczny. 2015. "Gender Gap through Time and Space: A Journey through Wikipedia Biographies and the 'wigi' Index." <http://arxiv.org/abs/1502.03086>.
- Lam, Shyong (Tony) K., Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl. 2011. "WP: Clubhouse: An Exploration of Wikipedia's Gender Imbalance." Pp. 1–10 in *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. New York: ACM.
- Laouenan, Morgane, Jean-Benoît Eyméoud, Olivier Gergaud, and Etienne Wasmer. 2018. "A Multi-language Database of Notable People (3000 BC-2015 AD)." Unpublished manuscript.
- Larivière, Chaoqun Ni, Yves Gingras, Blaise Cronin, and Cassidy R. Sugimoto. 2013. "Bibliometrics: Global Gender Disparities in Science." *Nature* 504:211–13.
- Lehrach, Hans, Don Diamond, John M. Wozney, and Helga Boedtker. 1977. "RNA Molecular Weight Determinations by Gel Electrophoresis under Denaturing Conditions, a Critical Reexamination." *Biochemistry* 16(21):4743–51.
- Long, J. Scott. 1992. "Measures of Sex Differences in Scientific Productivity." *Social Forces* 71(1):159–78.
- Luo, Wei, Julia Adams, and Hannah Brückner. 2018. "The Ladies Vanish? American Sociology and the Genealogy of Its Missing Women on Wikipedia." *Comparative Sociology* 17:519–56.
- Martin-Martin, Alberto, Enrique Orduna-Malea, Anne-Wil Harzing, and Emilio Delgado López-Cózar. 2017. "Can We Use Google Scholar to Identify Highly Cited Documents?" *Journal of Informetrics* 11(1):152–63.
- Merritt, Deborah Jones. 2000. "Scholarly Influence in a Diverse Legal Academy: Race, Sex, and Citation Counts." *The Journal of Legal Studies* 29:345–68.
- Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159(3810):56–63.
- Moss-Racusin, Corinne A., and Laurie A. Rudman. 2010. "Disruptions in Women's Self-promotion: The Backlash Avoidance Model." *Psychology of Women Quarterly* 34:186–202.
- Myers, Ben. 2016. "Where Are the Minority Professors?" *The Chronicle of Higher Education* (February 14). Retrieved January 10, 2019 (<https://www.chronicle.com/interactives/where-are-the-minority-professors>).
- Ngila, Dorothy, Nelius Boshoff, Frances Henry, Roseanne Diab, Shirley Malcom, and Jennifer Thomson. 2017. "Women's Representation in National Science Academies: An Unsettling Narrative." *South African Journal of Science* 113(7/8):1–7.
- Ray, Victor. 2018. "The Racial Politics of Citation." *Inside Higher Ed*. Retrieved August 12, 2018 (<https://www.insidehighered.com/advice/2018/04/27/racial-exclusions-scholarly-citations-opinion>).
- Reagle, Joseph, and Lauren Rhue. 2011. "Gender Bias in Wikipedia and Britannica." *International Journal of Communication* 5:1138–58.
- Rossiter, Margaret W. 1993. "The Matthew Matilda Effect in Science." *Social Studies of Science* 23(2):325–41.
- Rudman, Laurie, and Julie E. Phelan. 2008. "Backlash Effects for Disconfirming Gender Stereotypes in Organizations." *Research in Organizational Behavior* 28:61–79.
- Ruths, Derek, and Faiyaz Al Zamal. 2010. "A Method for the Automated, Reliable Retrieval of Publication-citation Records." *PloS One* 5(8):e12133.
- Sandberg, Cheryl. 2013. *Lean in: Women, Work, and the Will to Lead*. New York, NY: Alfred A. Knopf.
- Tripodi, Francesca. 2017. "The Silenced Minority—How Integrated Audiences Limit Participation Across Platforms." PhD dissertation, University of Virginia.
- Van den Besselaar, Peter, and Ulf Sandström. 2016. "Gender Differences in Research Performance and Its Impact on Careers: A Longitudinal Case Study." *Scientometrics* 106:143–62.
- Wagner, Claudia, Eduardo Graells-Garrido, David Garcia, and Filippo Menczer. 2016. "Women through the Glass Ceiling: Gender Asymmetries in Wikipedia." *EPJ Data Science Journal* 5(5).
- Wikimedia Foundation. 2011. "Wikipedia Editors Study: Results From the Editor Survey, April 2011." Retrieved January 10, 2019 (https://upload.wikimedia.org/wikipedia/commons/7/76/Editor_Survey_Report_-_April_2011.pdf).
- Wright, Erik Olin. 2011. "A Call to Duty: ASA and the Wikipedia Initiative." Retrieved March 16, 2018 (http://www.asanet.org/sites/default/files/savvy/footnotes/nov11/wikipedia_1111.html).
- Xie, Yu, and Kimberlee A. Shauman. 1998. "Sex Differences in Research Productivity: New Evidence about an Old Puzzle." *American Sociological Review* 63(6):847–70.

Author Biographies

Hannah Brückner works on a wide range of topics related to the life course, inequality, health, gender, and sexuality. Current research projects focus on the representation of academics and academic knowledge on Wikipedia and the impact of labor migration on gender inequality in Kerala (India). She has received an Andrew W. Mellon New Directions Fellowship and research grants from the

Robert Wood Johnson Foundation and the Volkswagen Foundation. Brückner is professor of social research and public policy at New York University Abu Dhabi, where she has served as associate dean of social sciences, vice provost for faculty diversity, program head, and chair of the Institutional Review Board.

Julia Adams teaches and conducts research in the areas of state building, social theory and knowledge, gender and family, early modern European politics, and colonialism and empire. Her current research focuses on large-scale forms of patrimonial politics, the historical sociology of agency relations, and gender, race, and the

representation of academic knowledge on Wikipedia and other digital platforms. Adams is professor of sociology and international and area studies and head of Grace Hopper College at Yale University. She also co-directs YaleCHESS (Center for Historical Enquiry and the Social Sciences).

Cambria Naslund graduated from NYU Abu Dhabi in 2015 with a degree in social research and public policy. She is currently pursuing a PhD in sociology at Princeton University. Her interests include computational social science, text analysis, and the sociologies of medicine, knowledge, and technology.