

Global Gender Differences in Wikipedia Readership

Isaac Johnson,¹ Florian Lemmerich,² Diego Sáez-Trumper,¹
Robert West,³ Markus Strohmaier,⁴ Leila Zia¹

¹Wikimedia Foundation,

²University of Passau,

³EPFL,

⁴RWTH Aachen University & GESIS – Leibniz Institute for the Social Sciences

isaac@wikimedia.org, florian.lemmerich@uni-passau.de, diego@wikimedia.org,

robert.west@epfl.ch, markus.strohmaier@cssh.rwth-aachen.de, leila@wikimedia.org

Abstract

Wikipedia represents the largest and most popular source of encyclopedic knowledge in the world, aiming to provide equal access to information. From a global online survey of 65,031 readers of Wikipedia and their corresponding reading logs, we present first evidence of gender differences in Wikipedia readership and how they manifest in records of user behavior. More specifically we report that (1) women are underrepresented among readers of Wikipedia, (2) women view fewer pages per reading session than men do, (3) men and women visit Wikipedia for similar reasons, and (4) men and women exhibit specific topical preferences. Our findings lay the foundation for identifying pathways toward knowledge equity in the usage of online encyclopedic knowledge.

Introduction

Equal access to encyclopedic knowledge represents a critical prerequisite for promoting equality and for an open and informed public at large. With almost 54 million articles written by roughly 500,000 monthly editors across more than 160 actively edited languages, Wikipedia is the most important source of encyclopedic knowledge worldwide, and one of the most important knowledge resources available on the internet. Every month, Wikipedia attracts users on more than 1.5 billion unique devices from across the globe, for a total of more than 15 billion monthly pageviews (Wikimedia Foundation 2020). Data about who is represented in this readership provides unique insights into the accessibility of encyclopedic knowledge on a global scale. Unlike the well-documented gender gap amongst Wikipedia contributors (Hill and Shaw 2013; Ford and Wajcman 2017; Antin et al. 2011; Lam et al. 2011; Collier and Bear 2012; Sichler and Prommer 2014), little is known about potential gender differences in *readership* and the similarities and differences between different sociodemographic populations of readers. This paper sets out to explore these open questions by conducting an in-depth study of gender differences in Wikipedia readership worldwide.

Approach. Building on past research (Glott, Schmidt, and Ghosh 2010; Zickuhr 2011; Protonotarios, Sarimpei, and

Otterbacher 2016; Kim, Sin, and Tsai 2014; Lim and Kwon 2010; Hinnosaar 2019; Selwyn and Gorard 2016) and surveys,¹ we provide insights into Wikipedia’s global readership and their reading behavior through 16 large-scale surveys of more than 65,031 Wikipedia readers across 14 different language editions that were conducted directly on the Wikipedia website. We link the readers’ responses with records² of their reading behavior on Wikipedia. Through this unique *combination of survey responses with actual navigation records*, we are able to identify sociodemographic differences in how Wikipedia is consumed.

Findings. We find that women are substantially underrepresented among readers of Wikipedia. Across 16 surveys, *men represent approximately two-thirds of Wikipedia readers on any given day*. Additionally, we observe that women view fewer pages per reading session than men do. However, we also find that on average, men and women visit Wikipedia for similar reasons. That is, the depth of knowledge that they seek, referred to as *information need* for the remainder of this paper, and their triggers for reading Wikipedia, referred to as *motivations*, are remarkably similar. Finally, men and women exhibit *specific topical preferences*. Readership of articles about sports, games, and mathematics is skewed towards men, while readership of articles about broadcasting, medicine, and entertainment is skewed towards women. We further observe evidence of *self-focus bias* (Hecht and Gergle 2009), i.e. that women tend to read relatively more biographies of women than men do, whereas men tend to read relatively more biographies of men than women do.

Implications. Our results indicate that globally, substantial barriers for women still exist in terms of knowledge equality.³ Combined with finding evidence for gender-specific reading behavior, this implies that popularity-based recommendations and rankings on the platform that do not take gender imbalance in readership (Wulczyn et al. 2016) into

¹https://meta.wikimedia.org/wiki/Category:Reader_surveys

²https://wikitech.wikimedia.org/wiki/Analytics/Data_Lake/Traffic/Webrequest#Current_Schema

³While our surveys allowed readers to self-describe their gender identity, we only received sufficient data from men and women to conduct robust analyses. See the Methods for more details.

account have the potential to further exacerbate existing gender imbalances on Wikipedia.

Data and Methods

We collected in-situ survey responses from 65,031 Wikipedia readers, alongside server logs of their click sessions during which the survey was completed.⁴ The survey was run worldwide in 14 Wikipedia languages (see Table 1) from June 26 to July 3, 2019 (with the exception of Polish Wikipedia, which was run from September 26 through October 31, 2019.) We selected the language editions with the following considerations in mind: diversity of language family, geographic diversity (as far as primary location of readers), and diversity of size of readership with the constraint that the language must receive sufficient pageviews to support the survey. In addition, we also included languages by requests of Wikipedia volunteers. For the globally spoken languages English and French, in addition to sampling users worldwide, we included a separate sampling procedure that specifically targeted Wikipedia readers in Africa (geolocated based on their IP addresses) to receive sufficient data to study potential regional differences in usage for these editions. This led to a total of 16 surveys across 14 languages that together represent 78% of the monthly pageviews across all language editions of Wikipedia.

Survey	# Resp. over 18	Countries with 500 responses or more
Arabic (ar)	7741	Saudi Arabia, Egypt, Iraq
German (de)	4144	Germany
English (en – Worldwide)	6181	USA, India
English (en – Africa)	8043	South Africa, Nigeria, Kenya, Egypt
Spanish (es)	11897	Spain, Mexico, Argentina, Columbia, Peru, Chile
Persian (fa)	7036	Iran
French (fr – Worldwide)	4401	France
French (fr – Africa)	3122	Morocco, Algeria
Hebrew (he)	586	Israel
Hungarian (hu)	1216	Hungary
Norwegian (no)	737	Norway
Polish (pl)	688	Poland
Romanian (ro)	1336	Romania
Russian (ru)	4565	Russia, Ukraine
Ukrainian (uk)	1148	Ukraine
Chinese (zh)	2190	Taiwan (Republic of China)

Table 1. Survey response counts and country breakdowns. For each survey, we provide the number of responses from individuals who indicated that they were over the age of 18 (what we analyze in this research). Additionally, for each survey, we provide the countries with at least 500 responses.

Survey Questions

The survey solicited information about the respondents’ demographics (age, gender, education, locale, native language) and their reasons for reading Wikipedia (motivation, information need, prior knowledge). Via server logs, the responses were enriched with information about the situational context (*e.g.*, geography, time of day) and the user’s behavior while reading Wikipedia (*e.g.*, session length, topics read, whether readers switched language editions while reading).

The survey questions were designed with the goal of balancing simplicity, privacy, and applicability to a global audience. Questions were adapted from prior, validated surveys where possible. In particular, we reused questions about motivation and information need from existing publications that targeted these topics specifically (Singer et al. 2017; Lemmerich et al. 2019), while for demographic questions we adapted questions from multiple sources including the International Social Survey Program (Edlund and Lindh 2019) and previous surveys on Wikipedia.

Utilizing the validated taxonomy of Wikipedia readership use-cases (Singer et al. 2017), we asked respondents three multiple-choice questions:

1. *I am reading this article because* (a) I have a work or school-related assignment; (b) I need to make a personal decision based on this topic (*e.g.*, buy a book, choose a travel destination); (c) I want to know more about a current event (*e.g.*, a soccer game, a recent earthquake, somebody’s death); (d) the topic was referenced in a piece of media (*e.g.*, TV, radio, article, film, book); (e) the topic came up in a conversation; (f) I am bored or randomly exploring Wikipedia for fun; (g) this topic is important to me and I want to learn more about it (*e.g.*, to learn about a culture). Users could select multiple answers.
2. *I am reading this article to* (a) look up a specific fact or to get a quick answer; (b) get an overview of the topic; (c) get an in-depth understanding of the topic.
3. *Prior to visiting this article I was* (a) already familiar with the topic; (b) not familiar with the topic, and I am learning about it for the first time.

Free-form answers were also allowed, but the vast majority of respondents chose from the pre-defined answers, suggesting that the taxonomy developed in the earlier study (Singer et al. 2017) remains comprehensive.

The five demographic questions were based on past surveys of readers and factors known to affect readership: gender, age, education level, locale (urban vs. rural), and native language. For this paper, we focused on gender. We applied best practices of allowing respondents to select from many identities or self-identify via inclusive language⁵ within the constraints of the Google Forms platform. We provided preset answers of “Man”, “Woman”, and “Prefer not to say” along with a free-text “Other” option. Although a number of respondents identified their gender as non-binary, the usage of “Other” was too low to perform language-specific statistical analyses (417 respondents across all surveys).

The five demographic questions were as follows:

⁴Privacy statement: https://foundation.wikimedia.org/wiki/2019_Wikipedia_Demographics_Survey_Privacy_Statement

⁵<https://www.morgan-klaus.com/gender-guidelines.html>

1. *What is your age?* (a) 18–24 years; (b) 25–29 years; (c) 30–39 years; (e) 40–49 years; (f) 50–59 years; (g) 60 years and older; (h) Prefer not to say.
2. *What is your gender?* (a) Woman; (b) Man; (c) Prefer not to say; (d) Other (free-form answer).
3. *How many years (full-time equivalent) have you been in formal education? Include all primary and secondary schooling, university and other post-secondary education, and full-time vocational training, but do not include repeated years. If you are currently in education, count the number of years you have completed so far.* (a) I have no formal schooling; (b) 1–6 years; (c) 7 years; (d) 8 years; (e) 9 years; (f) 10 years; (g) 11 years; (h) 12 years; (i) 13 years; (j) 14 years; (k) 15 years; (l) 16 years; (m) 17 years; (n) 18 years; (o) More than 18 years; (p) Prefer not to say.
4. *Would you describe the place where you live as* (a) A farm or home in the country; (b) A country village; (c) A small city or town; (d) The suburbs or outskirts of a big city; (e) A big city; (f) Prefer not to say.
5. *What is your native language?* (Choice from a list of Wikipedia languages in their native scripts with language codes and the survey language at the top of the list)
6. *What is your second native language?* (a) I do not have a second native language; (b) Other (free-form answer).

The survey was designed and piloted in English and then translated into the 13 other languages with the assistance of native speakers, comprised primarily of Wikipedia volunteers in these languages. Respondents could skip any of the demographic questions, although in practice response rates for those questions were always greater than 90% and in most cases above 95% of survey participants for all surveys. A link to the survey was displayed within Wikipedia articles as the reader was browsing (details below under “Survey Sampling”). This means that the three motivation and information need questions were answered in the context of a particular article the respondent was reading. For legal reasons, we additionally provided an initial screening question that removed individuals under the age of 18.

Survey Sampling

A prompt that asked users to participate in a survey in order to help us improve Wikipedia was shown using QuickSurveys extension⁶ as a box embedded toward the top of articles on the desktop and mobile versions of Wikipedia. Users who chose to participate were sent to a language-specific questionnaire hosted on Google Forms.

Sampling rates were set on a per-language basis and based on predictions on the number of daily pageviews expected in each language (ranging from 1:2 for Norwegian Wikipedia to 1:98 for English Wikipedia). For English and French Wikipedia, we applied two separate sampling procedures: one general random sample of global readership (referred to as “Worldwide”, this is how we sampled the other languages, too), and another one with filters that specifically sampled readers from African countries (referred to as “Africa”). The

first time a browser visited a Wikipedia language edition with an active survey, a random hash ID was generated that deterministically indicated whether the browser would see the survey or not. The hash ID was stored in the browser and remained there unless cookies were refreshed. If the survey was taken or dismissed, the hash was adjusted to indicate this and the survey was no longer shown on that browser. Thus, anyone using that particular browser would continue to see the survey on each Wikipedia article they viewed in that language until they either took the survey, dismissed it, or cleared their cookies. The survey was not shown on browsers with Do-Not-Track enabled. This sampling strategy was simple and preserved privacy (as all logic occurred on the client side), but had the limitation that an individual could potentially see and respond to the survey on multiple browsers (though in practice, we saw no evidence of this).

We also acknowledge that the identification of individual reader sessions cannot be guaranteed to always be 100% accurate—*e.g.*, there might be multiple individuals sharing one computer or a device’s IP address could change during a single session. Neither scenario would be detected by our approach, but, even if these situations occurred, there is no reason to believe this would introduce a strong systematic bias towards any particular gender identity in the results.

Survey Linking

When an individual responded to the survey, a unique code was passed through the survey that we use to directly link that individual’s response to the Wikipedia article that they were reading when they took the survey. We could then reconstruct that reader’s broader session of pageviews associated with the survey response, under the assumptions that the individual who took the survey was the only reader of Wikipedia on the browser in which the survey was taken in that period of time, that all of that reader’s pageviews came from a single device and browser, and that that device was associated with a single IP address. We define a reader session as consecutive pageviews with no more than one hour between subsequent pageviews based on past research (Halfaker et al. 2015). We limit our analyses to just the reader session in which the survey was taken to reduce the risk of violating our assumptions—*e.g.*, over longer time periods, we would expect mobile devices to change IP addresses.

Bias and Validity

Following the procedure from previous studies of Wikipedia readership through surveys (Singer et al. 2017; Lemmerich et al. 2019), we adjusted for potential biases in the survey responses to the best of our ability given the availability of features that may capture biases. To this end we used *inverse propensity weighting* (Austin 2011; Lunceford 2004) utilizing behavioral data extracted from the pageview logs. We compared the reading behavior of the survey respondents to the reading behavior of a representative sample of readers from that language edition. We learned weights that rebalance the survey-respondent population for a given language so that it has similar observed characteristics to the full reader population for that language edition. In particular, we controlled for the following covariates:

⁶<https://www.mediawiki.org/wiki/Extension:QuickSurveys>

- *contextual features*: day of week, time of day, country, continent
- *reader session features*: session time length, session access method, average time between pageviews, session referer class, where in session survey was taken, number of Wikipedia languages viewed, number of pageviews, reader logged-in, reader viewed Wikipedia’s Main Page
- *article demand features*: average and entropy of pageviews to articles read, average and entropy of number of languages each article is available in, pageviews to article where survey was taken
- *article topic features*: biography of a man, biography of a woman, article has latitude-longitude coordinates, article is for an event with a point-in-time, article topics
- *article quality features*: average and entropy of length of articles read, average infonoise (Warncke-Wang, Cosley, and Riedl 2013), average number of second and third-level headings, average number of templates in article, average number of references, average and entropy of number of internal links, average number of external links

With inverse propensity score weighting, each survey response gets assigned a weight that is the inverse of the predicted likelihood that that person would respond to the survey based on their reading behavior. For this prediction, we used a gradient boosting classifier as implemented in the scikit-learn Python library, using default parameters.⁷ To reduce the effect of strong outliers, the top 5% of weights were trimmed to the 95% threshold (Potter 1993), see also (Lee, Lessler, and Stuart 2011). The application of trimming did not change any of the main trends we discuss.

By analyzing the gradient boosting classifier and the resulting weights, we observed that the most noticeable effect of weighting was with respect to the number of articles viewed by a reader: readers who view more pages are also more likely to respond to the survey and thus are weighted lower in the results than readers who view a single page. In practice, this correction only shifted the results by at most 6%, with, *e.g.*, consistent shifts for our estimates of readers who are motivated by boredom (overrepresented in the raw results) and who identify as women (underrepresented in the raw results). These debiased results are used for all analyses with the exception of simple count-based results—*e.g.*, number of responses per country for a survey or total pageviews from survey respondents to a given article. Respective results for unweighted data can be found in the appendix, see also Section .

A number of additional robustness checks were implemented that validated the survey results. The three motivation-related questions in the survey were also asked of Wikipedia readers in several of the 14 languages in 2017 using very similar methods (Lemmerich et al. 2019). We saw for those questions that the results had largely remained stable between 2017 and 2019. For Russian and English, we also ran the survey for one week in June 2019 and for one month in September 2019 (not reported in this paper). Even

with the longer sampling time-frame, which would give infrequent readers more opportunities to see and respond to the survey, we saw nearly identical results (with small differences that were largely explained by seasonality).

Article Topics

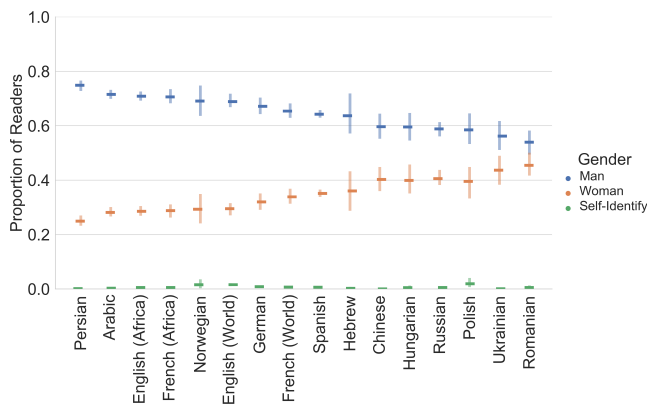
For our analyses, we compared the types of content that men and women read on Wikipedia. We faced the challenge that our surveys were deployed in 14 languages and the respondents of these surveys viewed articles in a total of over 100 languages over the course of their reading sessions. Traditional topic modeling would be computationally intensive and require extensive hand-labeling of topics in many languages to provide complete coverage of the dataset. To assign consistent topics across all of these languages, we instead relied on Wikidata, which contains language-independent structured data about the concepts covered in Wikipedia articles. Specifically, we employed two methods of categorizing the articles that the survey respondents viewed in our dataset: (1) we deterministically identified articles as biographies based on their associated Wikidata item, which specifies whether the article is about a person and, if so, what that person’s gender identity is, and (2) we trained a model to predict which of 44 high-level topics (such as *STEM.Medicine* or *Geography.Europe*) apply to any given Wikipedia article using a supervised fastText (Joulin et al. 2017) model trained on Wikidata attributes. The topics for this classification task were extracted from earlier research (Asthana and Halfaker 2018) on building a taxonomy for Wikipedia article topics that spreads across four high-level categories (STEM, Geography, History and Society, Culture), each of which have many sub-categories. Across these topics, we achieved a micro F1 score of 0.811 and macro F1 score of 0.643.

Statistical Analyses

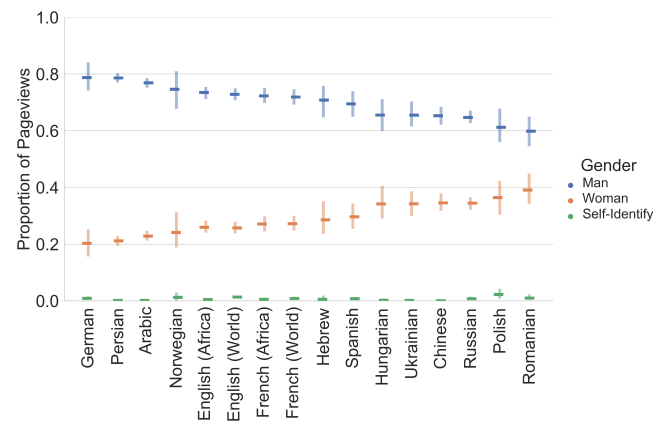
We present the results of simple comparison-based analyses (*e.g.*, whether the proportion of women aged 18–24 is significantly different than that of men aged 18–24) as they most directly reflect the composition of Wikipedia readers. Specifically, for each survey, we run a logistic regression⁸ for that survey’s responses. The dependent variable is a binary variable that indicates whether the respondent identified as a man. The independent variables are the categorical variables for topic, information need, country, age, education, locale, and native language. These regressions provide deeper insight into whether the correlations we observe can be explained by other observed factors. While in all cases we see support for our comparison-based conclusions, a lack of significance after controlling for other variables would in no way detract from the conclusion that women are significantly underrepresented in the global readership of Wikipedia. Instead it indicates that the reasons behind lower readership among women are likely complex, similar to the causes hypothesized for Wikipedia editors’ gender gap (Lam et al. 2011; Protonotarios, Sarimpei, and Otterbacher 2016; Hinnosaar 2019).

⁷<https://scikit-learn.org/0.19/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>

⁸Using the *glm* package in R with our debiased sample weights



(a) Gender of readers by language.



(b) Gender of pageviews by language.

Figure 1. Proportion of Wikipedia readers and pageviews by language and gender. (a), Proportion of readers who identified as men, women, or self-identified as non-binary. The gap in gender representation is the highest in Persian Wikipedia (75% men) and lowest in Romanian Wikipedia (54% men). No language edition had a higher proportion of women readers than men. (b), Proportion of pageviews by language from readers who identified as men, women, or self-identified as non-binary. The gap in gender representation in pageviews is the highest in German Wikipedia (79% men) and lowest in Romanian Wikipedia (60% men). However, even in the case of Romanian Wikipedia, this means that men generate 50% more pageviews than women.

Throughout the paper, we rely on 99% confidence intervals for deciding which results we consider to be significant. For the comparison-based analyses, we computed 99% confidence intervals through bootstrap resampling of the weighted survey responses with 400 iterations. Given that we have sixteen surveys and many variables under study, we face the challenge of multiple comparisons. We do not make Bonferroni corrections or other such adjustments but instead rely on the 99% confidence level and presence of 16 independent surveys to act as a natural check on false positives—namely, we do not report trends as significant unless we see them significant across at least 5 surveys along with no significant trends in the opposite direction from any of the surveys.

With the exception of our estimates of the proportion of women in worldwide readership and pageviews, we avoid directly combining our results—*e.g.*, a multi-level regression model that might produce a single significance value for a given covariate. For the former, it is clear that the survey results should be weighted by the proportion of pageviews each language produces as we seek a single number representative of the current state of global readership. As such the results for English Wikipedia (50% of pageviews) heavily influence the results while minor language editions such as the Romanian Wikipedia (0.2% of pageviews) only have little weight. For more advanced analyses, giving proportionally more weight to the English Wikipedia than to other language editions simply because it has more readers would obfuscate the findings for smaller language editions. Since the focus of this work was to investigate gender differences *worldwide*, we run our analyses separately for each survey and give each uniform weight in discussing the global trends that we observe.

Results

Next, we present our main empirical results.

Women Are Underrepresented Among Readers of Wikipedia

We estimate that globally across all surveyed language editions of Wikipedia, about two-thirds of readers on any given day are men.⁹ While Figure 1a shows substantial gender gap variation between languages (ranging from Romanian Wikipedia at 54% men to Persian Wikipedia at 75% men), it is notable that none of the language editions that we surveyed had a majority of women readers. This observation held true even for Norwegian Wikipedia readers (93% of whom are in Norway) for whom, informed by Norway’s ranking as second in the list of countries with the smallest Global Gender Gap (World Economic Forum 2020), we had hypothesized gender parity in readership.

We also found that within individual language editions whose readership is distributed more evenly across multiple countries, different countries can have very different proportions of women readers. For instance, women comprised 36% of readers in English Wikipedia in the United States, while they constituted only 24% of readers in India.

Although men and women respondents shared many of the same demographic attributes (see Table 2 for age, education, locality, and native language), we found evidence of greater gender parity amongst younger readers. In 6 of the 16 surveys, women were significantly more likely than men to report being of age 18–24, and in no survey men reported to be in that age group significantly more often than women. For other age groups, we observed the opposite: Men were

⁹Women comprise 32% of readers and people with non-binary identities comprise 1%.

significantly more likely to report being older: 25–29 (4 of 16 surveys), 30–39 (7 of 16 surveys), 40–49 (4 of 16 surveys), and 50–59 (4 of 16 surveys), and in no survey did women report to be in one of those age groups significantly more often than men.

Women View Fewer Pages per Reading Session than Men Do

By analyzing the pageview requests of survey respondents as available in Wikipedia’s web server logs, we found that, on average, men had longer reading sessions than women; i.e., men read more articles on average when visiting Wikipedia than women. This observation holds for all 16 survey populations, but the magnitude of the difference varies (see Table 3). The largest difference was observed for the German Wikipedia edition, in which the mean session length of men (3.94 articles viewed) was almost twice the mean session length of women (2.13 articles viewed). By contrast, in Polish Wikipedia, the difference was only 13% (2.29 articles viewed per session for men as compared to 2.02 articles viewed per session for women).

As a consequence of these longer sessions, we observe an even stronger gender gap when considering visits to individual articles (pageviews) as opposed to visits by distinct individuals (readers). We estimate that overall, men comprise

67% of readers but generate around 72% of pageviews in the surveyed Wikipedia editions on any given day. Women generate 27% and readers with non-binary identities generate 1% of pageviews. Figure 1b shows that results vary from German Wikipedia at 79% of pageviews by men to Romanian Wikipedia at 60%.

This overrepresentation of pageviews from readers who are men manifests strongly in the top-50 most-viewed articles viewed by survey respondents and aggregated across all survey languages during the week of the survey: we did not observe a single article that was viewed more often by women than men. The most-viewed article across all survey respondents was for the *Chernobyl disaster*. It received 2.5 million pageviews from all readers across all languages in the week of the survey with 305 pageviews from survey respondents, of which 68% came from men. The proportion of pageviews from men could range much higher, though, as with the *2019 Africa Cup of Nations*, with 243 pageviews from survey respondents, of which 85% came from men. The lowest proportion of pageviews from men in these highly-viewed articles was for the article about the actor *Mehdi Hashemi*, where only 51% of the 155 survey respondent pageviews were from men.

Men and Women Visit Wikipedia for Similar Reasons

In the survey, we also asked Wikipedia readers across language editions about their information need (i.e., the depth of information they sought when on Wikipedia: a specific fact or a quick answer, an overview, or an in-depth read) as well as their motivation for visiting the site. We then

Demographic	Skew Men	Skew Women	No Diff.
Age 18–24	0	6	10
Age 25–29	4	0	12
Age 30–39	7	0	9
Age 40–49	4	0	12
Age 50–59	4	0	12
Age 60+	3	2	11
Education 0–11 years	2	2	12
Education 12 years	2	3	11
Education 13–16 years	0	1	15
Education 17–18 years	0	0	16
Education 18+ years	1	0	15
Locale City	1	2	13
Locale Suburbs	1	0	15
Locale Town	1	0	15
Locale Village	3	1	12
Locale Rural Area	0	0	16
Lang. Monoling. / Native	1	4	11
Lang. Multiling. / Native	0	3	13
Lang. Non-native	4	0	12

Table 2. Language count of significant differences between genders for demographic attributes across surveys. We look at how many of the 16 surveys had a significant difference between men and women for how old they were. For example, in 6 of the 16 surveys, women were significantly more likely than men to report being age 18–24 (and there was no statistically significant difference between men and women in the other 10 surveys). We see trends with varying levels of evidence that men are more likely to report being over the age of 24. Significance determined based on bootstrap resampling and 99% confidence interval.

Survey	# Requests (Men)	# Requests (Women)
Arabic	2.47 [2.35, 2.61]	1.86 [1.75, 2.01]
German	3.94 [3.13, 5.51]	2.13 [1.92, 2.39]
English (World)	2.86 [2.71, 3.05]	2.36 [2.17, 2.60]
English (Africa)	2.42 [2.30, 2.54]	2.12 [2.00, 2.34]
Spanish	2.79 [2.53, 3.26]	2.18 [1.96, 2.67]
Persian	2.71 [2.58, 2.88]	2.19 [2.03, 2.40]
French (World)	2.83 [2.60, 3.07]	2.07 [1.89, 2.35]
French (Africa)	2.06 [1.94, 2.20]	1.90 [1.77, 2.06]
Hebrew	2.23 [1.93, 2.54]	1.60 [1.41, 1.87]
Hungarian	2.36 [2.13, 2.71]	1.84 [1.60, 2.16]
Norwegian	2.43 [2.07, 3.29]	1.85 [1.57, 2.20]
Polish	2.29 [2.07, 2.59]	2.02 [1.73, 2.36]
Romanian	2.30 [2.01, 2.80]	1.78 [1.64, 1.97]
Russian	2.65 [2.50, 2.83]	2.05 [1.94, 2.19]
Ukrainian	2.77 [2.41, 3.29]	1.86 [1.67, 2.11]
Chinese	3.07 [2.80, 3.36]	2.41 [2.18, 2.81]

Table 3. Average number of pageviews for men and women readers across surveys. We look at how many of the 16 surveys had a significant difference between men and women for the average number of pageviews in their reading session. Across all surveys, we see that men view significantly more articles per reading session than women do. Significance determined based on bootstrap resampling and 99% confidence intervals (shown in parentheses).

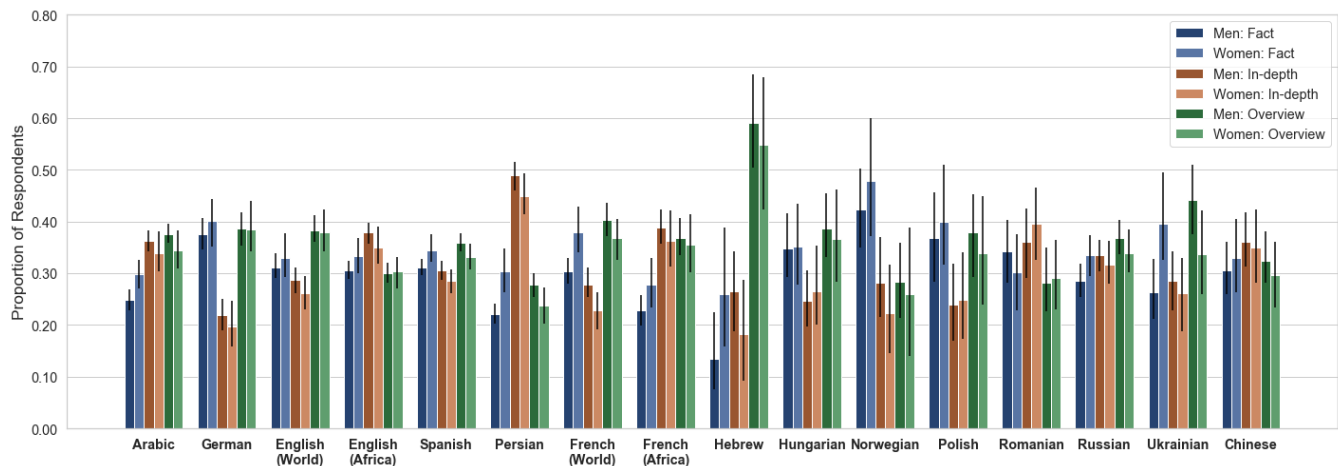


Figure 2. Information needs per gender across Wikipedia languages. We report with 99% confidence intervals the proportion of men and women for each category of information need: (i) facts, (ii) overview and (iii) in-depth information. The data shows that men and women have overall very similar information needs.

analyzed whether there are differences between men and women with respect to information need and motivation.

At a high level, men and women reported similar information needs (see Figure 2). For 9 of the 16 surveys, there were no significant differences between men and women for any reported information need. For the other seven, women reported looking for a fact significantly more than men. Differences between languages, however, were often much larger than differences between men and women; *e.g.*, in Arabic Wikipedia, women reported looking for facts 30% of the time, while men reported looking for facts 25% of the time; by contrast in German Wikipedia, men looked for a fact 40% and women 38% of the time.

Considering motivations to read Wikipedia (no detailed results shown here), the same high-level trends were seen amongst men and women (again with a few exceptions). We saw little evidence of consistent gender differences in any of the following motivations: conversation (2/16 surveys with significant differences), media (3/16), intrinsic learning (5/16), or help making a personal decision (2/16). In 12 of the 16 survey populations, however, men were more likely to report boredom or randomly reading Wikipedia for fun as their motivation for browsing Wikipedia. Men also were more likely to be motivated by a current event for checking Wikipedia in 7 of the 16 survey populations. Finally, women, reported work or school as a motivation significantly more often than men in 7 of the 16 survey populations. There were no significant differences in the respective other direction for other surveys with respect to boredom, current events, or work and school.

Men and Women Exhibit Specific Topical Preferences

While many topics on Wikipedia were read equally frequently by men and women, we found that across the 16 surveys, there were both topics that were read more consistently by men (*e.g.*, sports or technology) and topics that

were read more consistently by women (*e.g.*, broadcasting or medicine). In Figure 3 we visualize the gender skew for each topic in each language edition.

However, it is important to note that even for topics such as broadcasting (*e.g.*, television shows) or medicine, which women were more likely to read than men, men still generated the majority of pageviews. For instance, articles about the Chernobyl miniseries were the 13th-most-read articles by the survey population (122 times across all of the languages). Although the articles are categorized as broadcasting and women were more likely to read about the Chernobyl miniseries than men, 68% of the pageviews still came from men. This can be explained by our earlier observation that men have a higher general frequency of reading Wikipedia and longer reading sessions.

An interesting observation with respect to topical interest is that there was substantial self-focus in the most popular topic on Wikipedia: biographies (about 35% of pageviews). Figure 3 reveals that biographies (Culture.Biography) are balanced in how often they were viewed by men and women overall. A different picture emerges when considering the gender of the person who is described in the viewed biography. Figure 4 demonstrates clear self-focus bias: men were more likely to read biographies of men than women were (7 of 16 survey populations), whereas women were more likely to read biographies of women than men were (7 of 16 survey populations, with the other surveys showing no significant differences in both cases).

Discussion and Related Work

In this section, we present the implications of this study, methodological limitations, and related work.

Awareness vs. readership. An important distinction should be made about the readership measured in these surveys: while the results of this study show that the readers of many Wikipedia language editions on any given day are predominantly men, this does not necessarily imply that fewer

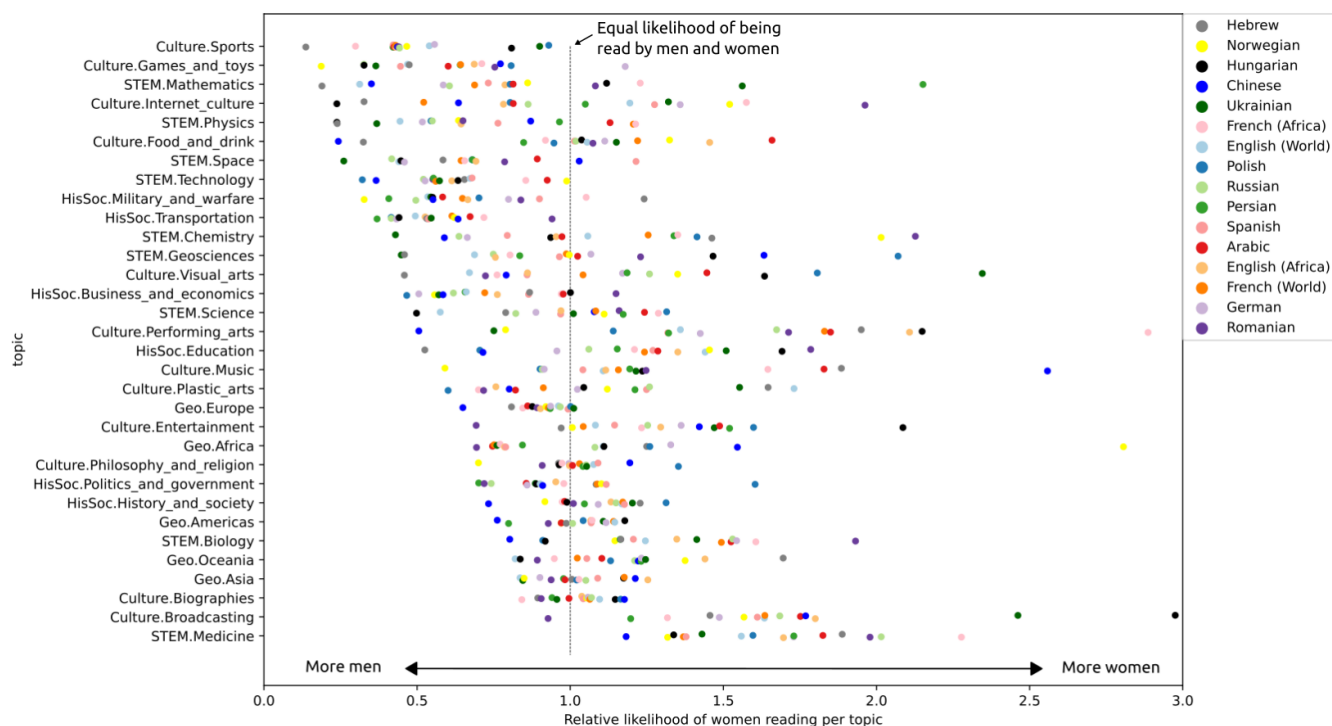


Figure 3. Topical interest in Wikipedia by gender and language. The relative frequency of reading about various topics on Wikipedia for women and men. $P(\text{topic}|\text{woman})/P(\text{topic}|\text{man})$. 1 indicates that men and women are equally likely to view a topic, 1.5 indicates that women are 50% more likely to read about the topic, and 0.66 indicates that men are 50% more likely to read about the topic. For all topics, women still generate fewer pageviews than men given the overall imbalance of pageviews from men. In topics such as Sports and Games, the share of women interest (on average) is low for all languages, with exception of women reading about *Sports* in the German Wikipedia. Topics like *Medicine* receive relatively more interest from women. In the majority topics such as *Philosophy and Religion* or most geographic topics, the attention between men and women either has no significant differences in any survey or just a few significant differences but mixed in their direction.

women read Wikipedia. In fact, a number of surveys have found that women are just as likely as men in many regions to identify as readers of Wikipedia, though less likely to have read Wikipedia in the recent past (Glott, Schmidt, and Ghosh 2010; Zickuhr 2011; Protonotarios, Sarimpei, and Otterbacher 2016); *e.g.*, “do you read Wikipedia?” as compared to “did you read Wikipedia yesterday?”. Together, these surveys and past surveys indicate that, while women may be equally aware of Wikipedia, on average they visit less frequently and, as we have shown, read fewer pages when they do visit. The reader behavior analyses we presented here further offer new insights into how these gaps might arise and manifest in research and tools that are built with reader data. We hope that this data stands as a baseline and motivation for increased efforts to address these gaps.

Content and readership diversity. Although the focus of our surveys was on measuring the readership gender gap as opposed to investigating its underlying causes, our comparison of reading behavior between men and women suggests possible mechanisms. In particular, our research raises questions about the impact of content gaps on who reads Wikipedia. The gender gap in Wikipedia *content* is well-documented (Graells-Garrido, Lalmas, and Menczer 2015;

Lam et al. 2011; Klein et al. 2016; Reagle and Rhue 2011; Wagner et al. 2015); *e.g.*, less than 20% of biographies in English Wikipedia are about women (Wade and Zaringham 2018). Prior research has suggested that women receive less value from Wikipedia (Lim and Kwon 2010; Garrison 2015), likely in part due to these content gaps. We did not see any indications in our data that women tended to read lower-quality content on Wikipedia. In particular, there were no consistent differences in the average length of articles, number of headings, or other features that are related to article quality (Warncke-Wang, Cosley, and Riedl 2013). We observed, however, that women showed more interest than men in topics such as biographies of women, medicine, and broadcasting. The differences in topics of interest can be expected in principle given previous studies of general media consumption (McMahon 2002; An and Kwak 2016; Shearer and Matsa 2018), however, missing, low-quality, or otherwise biased Wikipedia content specifically in these areas can have an outsized impact on women readers. This finding highlights the importance of work conducted by Wikipedia volunteers to increase content about women in Wikipedia (Wade and Zaringham 2018; Halfaker 2017) as well as the need for more frequent mon-

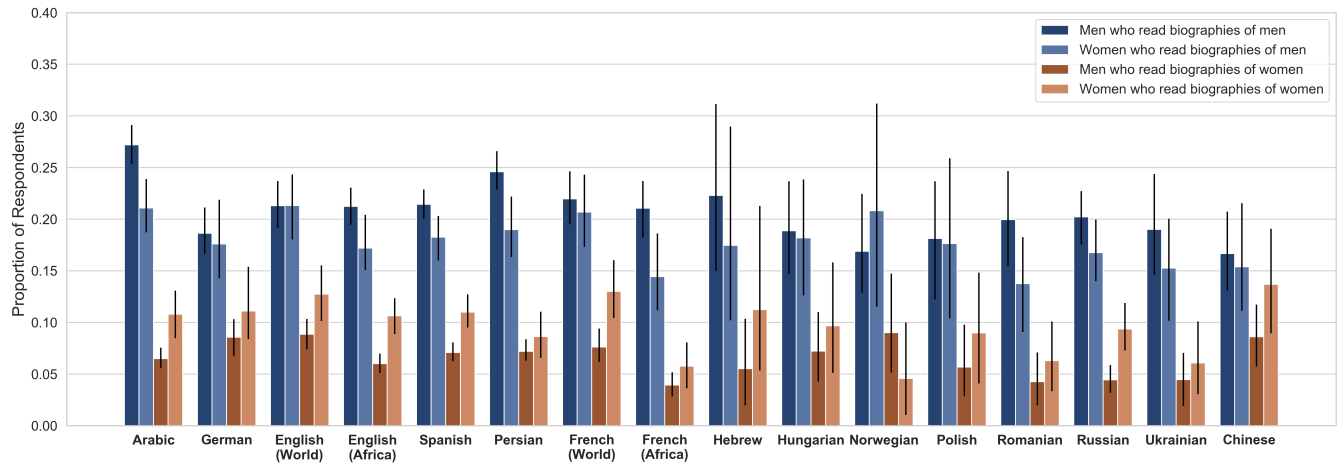


Figure 4. Interest in Wikipedia biographies about men and women by reader gender. The figure shows the proportion of men, resp. women, viewing at least one article with a biography about a man/woman. We can see that in most languages women read articles about women comparatively more often.

itoring of the gender differences in Wikipedia’s readership through multi-lingual large-scale surveys.

We further note that closing content gaps is not a panacea as evidenced by prior research on Welsh Wikipedia, where a majority of the biographies are about women (Lubbock), a majority of Welsh speakers are women,¹⁰ but readership is still heavily skewed towards men (Nevell, Galvez, and Owain 2017).

Editorship and readership diversity. Wikipedia readership gaps relate directly to the gender gaps among Wikipedia editors. The pipeline of participation suggests that the gender gap observed in contributors is in fact a function of gaps that appear in awareness of Wikipedia, readership of Wikipedia, awareness of the ability to edit Wikipedia, and only then retention of Wikipedia editors (Shaw and Hargitai 2018). Our surveys provide global data about the state of gender gaps in readership. If unaddressed, there will continue to be large gender gaps amongst Wikipedia contributors (Hill and Shaw 2013; Ford and Wajcman 2017; Antin et al. 2011; Lam et al. 2011; Collier and Bear 2012; Sichler and Prommer 2014), which can reinforce associated content gaps due to self-focus bias (Hecht and Gergle 2010; Das, Hecht, and Gergle 2019).

Pageviews and prioritization. In Wikipedia, the gender gap amongst readers is smaller than the gender gap observed amongst editors with estimates showing that only 9% of Wikimedia project editors identify as women.¹¹ Therefore even in the presence of the results of this study, the interests

of readers can be a valuable signal for guiding the prioritization of content to be more representative of what people (and not just existing readers) need or want to know.

The above being said, we encourage researchers, developers, decision makers such as those in the Wikimedia Foundation and other Wikimedia affiliates, and editors who use Wikipedia pageview data in their work to exercise caution when doing so in light of the findings of this study and to assure that their usage of Wikipedia’s pageview data does not reinforce the already existing biases in Wikipedia (and through that the broader Web). For example, we encourage editors that rank red links—Wikipedia article pages that do not exist and should be created—by their pageviews to also consider other features that can complement the raw pageview data used to prioritize missing content. Recommender systems such as (Wulczyn et al. 2016; Cosley et al. 2007) can also benefit from the evaluation of the role of pageviews to assure that the content recommended for creation is inclusive of all genders.

Beyond binary. Across all the surveys, we had 417 individuals who identified themselves as non-binary (Ukrainian with 2 responses had the smallest number and Spanish with 88 the largest). While the numbers from this group were small in this study, we reported the results associated with the group so long as we could confidently do so. We encourage future research to follow the practices laid out in a Wikimedia report on advancing gender equity (Stephenson-Goodknight et al. 2018) to help expand the debate around the gender gap in Wikipedia beyond a binary.

Methodological considerations. As all survey-based studies, our work might be subject to response biases, i.e., some groups will be overrepresented in the survey responses due to higher participation rates. As previous survey studies on Wikipedia (Singer et al. 2017; Lemmerich et al. 2019),

¹⁰<https://stats.wales.gov.wales/Catalogue/Welsh-Language/Annual-Population-Survey-Welsh-Language/welsh-skills-by-age-sex>

¹¹https://meta.wikimedia.org/wiki/Community_Insights/2018_Report/Diversity_of_contributors_on_the_Wikimedia_projects_seems_to_remain_unchanged.

we approached this issue by applying inverse propensity weighting as a state-of-the-art technique for debiasing results. Inverse propensity weighting for survey debiasing typically is applied with respect to demographic attributes, for which the distribution is known in the general population. However, for the Wikipedia readership these attributes are so far unknown – learning more about the readership demographics was a reason for the survey conducted as part of this study. Instead, we perform inverse propensity score weighting via user characteristics we extract from server logs and aim to adjust according to those properties. Yet, it cannot be guaranteed that our procedure corrects for all potential biases. Furthermore, it is theoretically possible that existing biases in the survey even get reinforced by our attempts to debias the data. Therefore, we include the main analytical results for non-debiased data in the appendix of this work (See Figures 5 and 6). Observing qualitatively the same outcomes for this data hints at the robustness of the results with respect to potential issues through biased survey responses. In addition, our survey results might also be influenced by other types of biases such as social desirability and/or translation biases. We refer to previous literature (Lemmerich et al. 2019) on the topic for a more in-depth discussion.

Conclusions

In this paper, we thoroughly investigated gender differences in the Wikipedia readership across the world by combining large-scale surveys and log data analysis. We found a consistent gender gap amongst Wikipedia readers in that women are underrepresented and have shorter sessions, have similar motivations and information needs as men, but different topical interests. In particular, we observe a tendency to self-focus, i.e., women tend to read biographies about women more often than men.

These results indicate that there remains large gaps in usage of Wikipedia and likely barriers to equal access to encyclopedic knowledge across gender identities. Investigating potential causes for this gender gap in readership will be a challenging and crucial task for future research. In the future, we aim to extend our analysis of specific topical interest to other demographic groups. Furthermore, we are interested to investigate the feedback cycle of the gender gap in content, editorship, and readership. Rerunning surveys as the one presented in this paper, could also allow for longitudinal studies of the gender gap in Wikipedia readership, a key step in assessing how well we are collectively doing in addressing Wikipedia’s gender gaps.

Acknowledgments

We thank Bahodir Mansurov, Wikimedia Foundation (WMF) Readers Web and Legal teams, and the following volunteers for helping to make the surveys happen: Abbad, Amire80, Yury Bulka, Adélaïde Calais WMFr, Kaganer, Liang-chih ShangKuan, Jon Harald Søby, Strainu, Tgr, Nettrom. We thank Clemens Lechner (GESIS - Leibniz Institute for the Social Sciences) for providing advice on the selection and harmonization of survey questions. Robert West is a Wikimedia Foundation Research Fellow. His lab

was partly supported by Swiss National Science Foundation grant 200021_185043 and by gifts from Google, Facebook, and Microsoft.

Appendix

This appendix contains figures that provide additional supporting evidence for claims made in the paper.

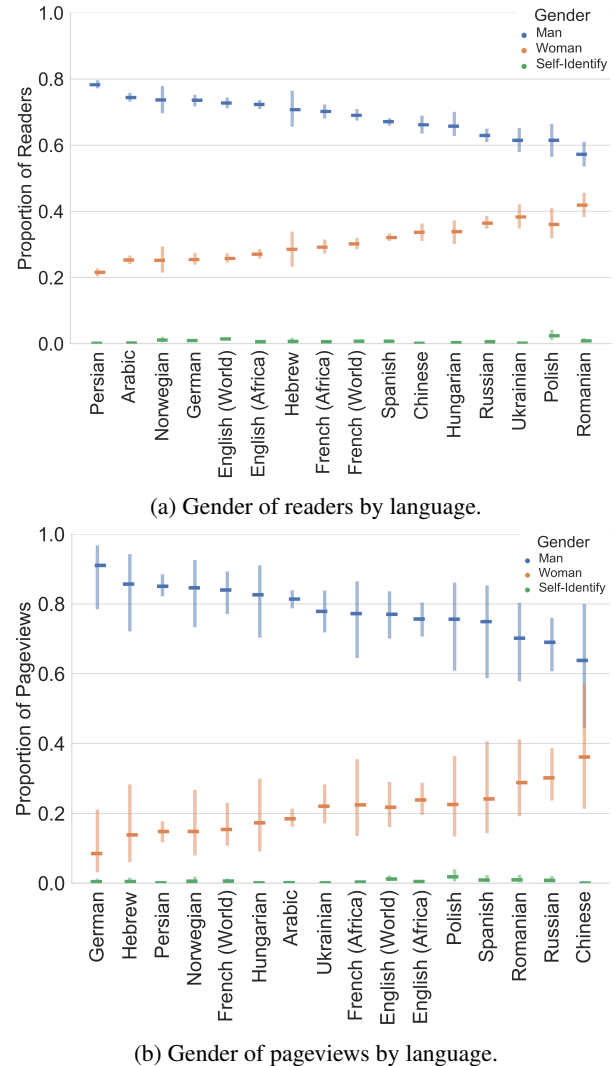


Figure 5. Proportion of Wikipedia readers and pageviews by language and gender with non-debiased data, i.e., no inverse propensity score weighting. We refer to the caption of Figure 1 for details.

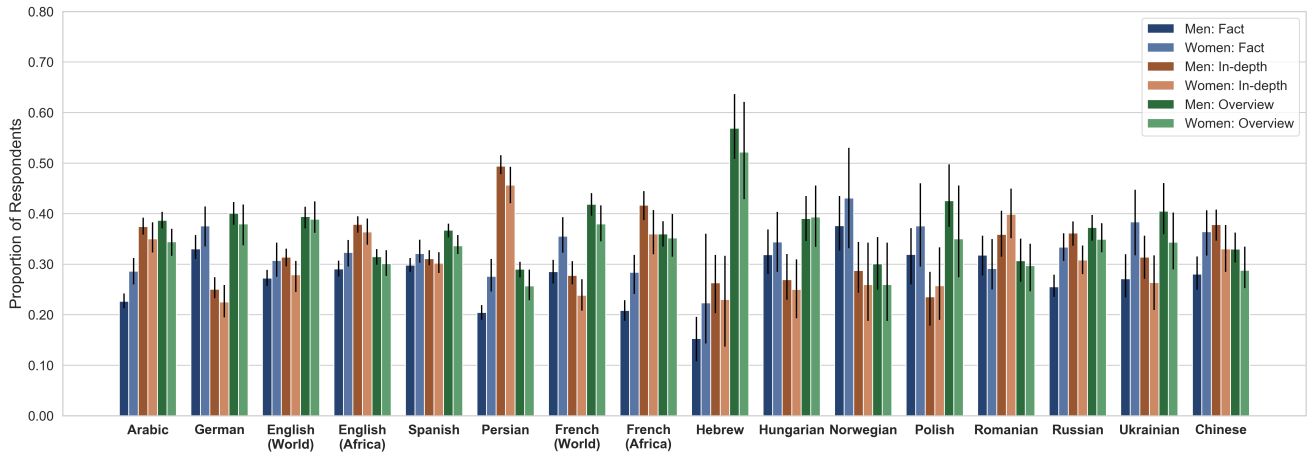


Figure 6. Information needs per gender across Wikipedia languages with non-debiased data, i.e., no inverse propensity score weighting. We refer to the caption of Figure 2 for details.

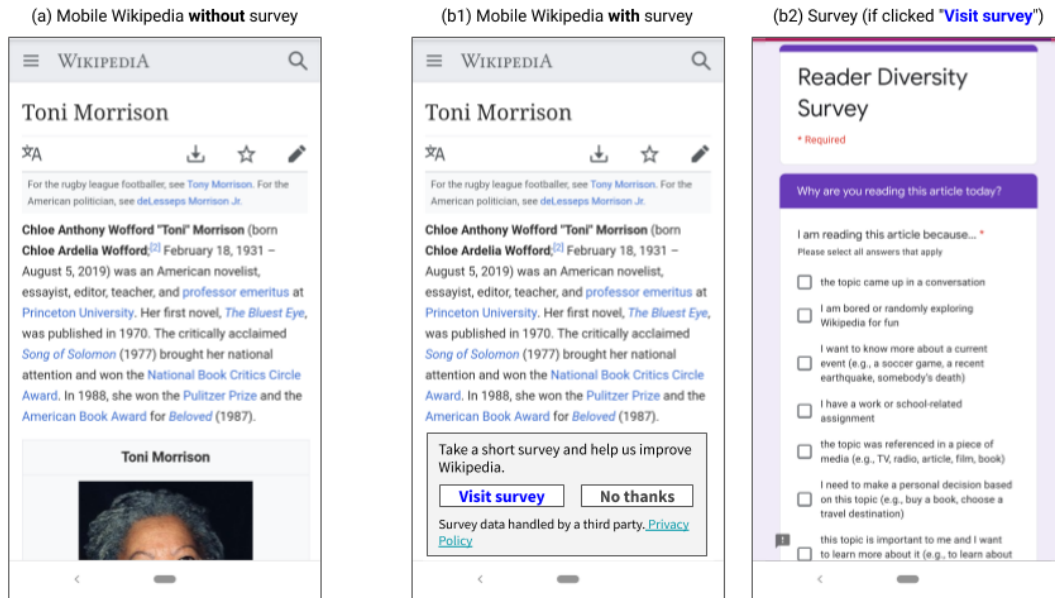


Figure 7. Survey participant recruitment on Wikipedia. Panel (a) shows how an article is normally displayed on mobile. Panel (b1) shows where the survey invitation would be inserted within the article. The survey invitation behaves similar to the article content and the reader can scroll past the survey. Panel (b2) shows a page from the survey if the reader clicked “Visit survey” (the initial page asks if the reader is above the age of 18). Display for desktop Wikipedia not shown but similar.

References

- An, J., and Kwak, H. 2016. Multidimensional analysis of gender and age differences in news consumption. In *Computation+ Journalism Symposium*.
- Antin, J.; Yee, R.; Cheshire, C.; and Nov, O. 2011. Gender differences in wikipedia editing. In *Proc. of the 7th Intl. Symposium on Wikis and Open Collaboration*, 11–14.
- Asthana, S., and Halfaker, A. 2018. With few eyes, all hoaxes are deep. *Proc. of the ACM on Human-Computer Interaction* 2:1–18.
- Austin, P. C. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* 46(3):399–424.
- Collier, B., and Bear, J. 2012. Conflict, confidence, or criticism: An empirical examination of the gender gap in wikipedia. In *Proc. of Conf. on Computer Supported Cooperative Work*, 383–392.
- Cosley, D.; Frankowski, D.; Terveen, L.; and Riedl, J. 2007. Sug-gestbot: using intelligent task routing to help people find work in wikipedia. In *Proc. of the 12th Intl. Conf. on Intelligent User In-terfaces*, 32–41.
- Das, M.; Hecht, B.; and Gergle, D. 2019. The gendered geography of contributions to openstreetmap: Complexities in self-focus bias. In *Proc. of the 2019 CHI Conf. on Human Factors in Computing*

Systems, 1–14.

Edlund, J., and Lindh, A. 2019. The issp 2016 role of government module: Content, coverage, and history. *Intl. J. of Sociology* 49(2):99–109.

Ford, H., and Wajcman, J. 2017. ‘anyone can edit’, not everyone does: Wikipedia’s infrastructure and the gender gap. *Social studies of science* 47(4):511–527.

Garrison, J. C. 2015. Getting a “quick fix”: First-year college students’ use of wikipedia. *First Monday* 20(10).

Glott, R.; Schmidt, P.; and Ghosh, R. 2010. Wikipedia survey—overview of results. *United Nations University: Collaborative Creativity Group* 8:1158–1178.

Graells-Garrido, E.; Lalmas, M.; and Menczer, F. 2015. First women, second sex: Gender bias in wikipedia. In *Proc. of the 26th ACM Conf. on Hypertext & Social Media*, 165–174.

Halfaker, A.; Keyes, O.; Kluver, D.; Thebault-Spieker, J.; Nguyen, T.; Shores, K.; Uduwage, A.; and Warncke-Wang, M. 2015. User session identification based on strong regularities in inter-activity time. In *Proc. of the 24th Intl. Conf. on World Wide Web*, 410–418.

Halfaker, A. 2017. Interpolating quality dynamics in wikipedia and demonstrating the keilana effect. In *Proc. of the 13th Intl. Symposium on Open Collaboration*, 1–9.

Hecht, B., and Gergle, D. 2009. Measuring self-focus bias in community-maintained knowledge repositories. In *Proc. of the fourth intl. conference on Communities and technologies*, 11–20.

Hecht, B., and Gergle, D. 2010. The tower of babel meets web 2.0: user-generated content and its applications in a multilingual context. In *Proc. of the SIGCHI Conf. on human factors in computing systems*, 291–300.

Hill, B. M., and Shaw, A. 2013. The wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PloS one* 8(6).

Hinnosaar, M. 2019. Gender inequality in new media: Evidence from wikipedia. *J. of Economic Behavior & Organization* 163:262–276.

Joulin, A.; Grave, E.; Bojanowski, P.; and Mikolov, T. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 427–431. Association for Computational Linguistics.

Kim, K.-S.; Sin, S.-C. J.; and Tsai, T.-I. 2014. Individual differences in social media use for information seeking. *The J. of Academic Librarianship* 40(2):171–178.

Klein, M.; Gupta, H.; Rai, V.; Konieczny, P.; and Zhu, H. 2016. Monitoring the gender gap with wikidata human gender indicators. In *Proc. of the 12th Intl. Symposium on Open Collaboration*, 1–9.

Lam, S. T. K.; Uduwage, A.; Dong, Z.; Sen, S.; Musicant, D. R.; Terveen, L.; and Riedl, J. 2011. Wp: clubhouse? an exploration of wikipedia’s gender imbalance. In *Proc. of the 7th Intl. symposium on Wikis and open collaboration*, 1–10.

Lee, B. K.; Lessler, J.; and Stuart, E. A. 2011. Weight trimming and propensity score weighting. *PloS one* 6(3).

Lemma, F.; Sáez-Trumper, D.; West, R.; and Zia, L. 2019. Why the world reads wikipedia: Beyond english speakers. In *Proc. of the Intl. Conf. on Web Search and Data Mining*, 618–626.

Lim, S., and Kwon, N. 2010. Gender differences in information behavior concerning wikipedia, an unorthodox information source? *Library & information science research* 32(3):212–220.

Lubbock, J. The Welsh Gender Equilibrium: Welsh becomes the biggest language Wikipedia to achieve gender balance!

<https://blog.wikimedia.org/2016/12/the-welsh-gender-equilibrium-welsh-becomes-the-biggest-language-wikipedia-to-achieve-gender-balance/>. Accessed: 2020-06-03.

Lunceford, J. K.; Davidian, M. 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine* 23(19):2937–2960.

McMahon, L. 2002. Understanding the dynamics of youth readership. *Young Consumers: Insight and Ideas for Responsible Marketers* 3(2):53–61.

Nevell, R.; Galvez, E.; and Owain, R. 2017. Readership of Welsh Wikipedia. https://meta.wikimedia.org/wiki/Research:Readership_of_Welsh_Wikipedia. Accessed: 2021-04-16.

Potter, F. J. 1993. The effect of weight trimming on nonlinear survey estimates. In *Proc. of the American Statistical Association, Section on Survey Research Methods*, volume 758763.

Protonotarios, I.; Sarimpei, V.; and Otterbacher, J. 2016. Similar gaps, different origins? women readers and editors at greek wikipedia. In *Tenth Intl. AAAI Conf. on Web and Social Media*.

Reagle, J., and Rhue, L. 2011. Gender bias in wikipedia and britannica. *Intl. J. of Communication* 5:21.

Selwyn, N., and Gorard, S. 2016. Students’ use of wikipedia as an academic resource—patterns of use and perceptions of usefulness. *The Internet and Higher Education* 28:28–34.

Shaw, A., and Hargittai, E. 2018. The pipeline of online participation inequalities: The case of wikipedia editing. *J. of communication* 68(1):143–168.

Shearer, E., and Matsa, K. E. 2018. News use across social media platforms 2018. Accessed: 2021-04-16.

Sichler, A., and Prommer, E. 2014. Gender differences within the german-language wikipedia. *ESSACHESS—Journal for Communication Studies* 7(2 (14)):77–93.

Singer, P.; Lemmerich, F.; West, R.; Zia, L.; Wulczyn, E.; Strohmaier, M.; and Leskovec, J. 2017. Why we read wikipedia. In *Intl. Conf. on World Wide Web*.

Stephenson-Goodknight, R.; Wang, A.; Johnson, M.; Houston, S.; Cruz, M.; and Reilly, L. 2018. Gender equity report 2018. Accessed: 2021-04-16.

Wade, J., and Zaringhalam, M. 2018. Why we’re editing women scientists onto wikipedia. *Nature* 14.

Wagner, C.; Garcia, D.; Jadidi, M.; and Strohmaier, M. 2015. It’s a man’s wikipedia? assessing gender inequality in an online encyclopedia. In *Ninth Intl. AAAI Conf. on web and social media*.

Warncke-Wang, M.; Cosley, D.; and Riedl, J. 2013. Tell me more: an actionable quality model for wikipedia. In *Proc. of the 9th Intl. Symposium on Open Collaboration*, 1–10.

Wikimedia Foundation. 2020. Wikimedia statistics - monthly overview. <https://stats.wikimedia.org/#/all-wikipedia-projects>. Accessed: 2021-04-16.

World Economic Forum. 2020. Global gender gap report. http://www3.weforum.org/docs/WEF_GGGR_2020.pdf. Accessed: 2021-04-16.

Wulczyn, E.; West, R.; Zia, L.; and Leskovec, J. 2016. Growing wikipedia across languages via recommendation. In *Proc. of the 25th Intl. Conf. on World Wide Web*, 975–985.

Zickuhr, K. 2011. *Wikipedia, past and present*. Pew Internet & American Life Project.