THE SWEDISH WIKIPEDIA GENDER GAP

Björn Helgeson
bhelg@kth.se

Swedish Title: ”Fördelningen kvinnor och män på svenska Wikipedia”

Industry Supervisor: Sara Mörtsell,
Wikimedia Sverige

Thesis Supervisor: Christopher Rosenqvist,
Stockholm School of Economics

Examiner: Dr. Haibo Li,
Royal Institute of Technology

Abstract

The proportion of women editors on the English language Wikipedia has for years been known to be very low. The purpose of this thesis is to see if this gender gap exists on the Swedish language Wikipedia as well, and investigate the reasons behind it.

To do this, three methods are used. Firstly a literature review is conducted, looking at women in computing and how Wikipedia works and how it was founded. Secondly, user behavior and activity-levels are measured through means of a database analysis of editors and edits. And thirdly, a survey is distributed, aimed at both readers and editors of Swedish Wikipedia, gathering some 2700 respondents.

The results indicate that there is indeed a big disproportion, and that only between 13-19% of editors are women. The findings did not indicate readers of the encyclopedia having any strong negative preconceptions about Wikipedia or its community. However when looking at reasons for not contributing, women were significantly more likely to perceive themselves as not competent enough to edit. Computer skills were found to be an important factor for trying out editing in the first place, and Wikipedia's connection to a male-dominated computing/programming culture is put forth as a reason for the resilience of the gender gap. The difference in men's and women's communication styles in relation to the climate Wikipedia's policies and guidelines is also discussed.

Sammanfattning

Andelen kvinnliga redigerare på engelskspråkiga Wikipedia har visats vara väldigt låg. Syftet med detta arbetet är att undersöka om andelen ser likadan ut på den Svenskspråkiga siten också, samt undersöka de bakomliggande orsakerna.

För att göra detta används tre metoder. Först görs en literaturstudie som behandlar kvinnor inom programmering och hur Wikipedia fungerar och dess grundande. Därefter mäts användarbeteende och aktivitetsnivåer genom en databasanalys på redigerare och redigeringar. Slutligen distribuerades en webb-enkät riktad till både läsare och redigerare av svenskspråkiga Wikipedia, med runt 2700 svaranden.

Resultaten visar att det finns en stor snedfördelning och att endast mellan 13-19% av redigerare är kvinnor. Resultaten påvisar inte några särskilda negativa uppfattningar hos läsare om Wikipedia eller dess gemenskap. Däremot uppgav kvinnor i signifikant högre utsträckning att en viktig anledning till att de inte bidrog till encyklopedin var att de inte upplevde sig tillräckligt kompetenta. Datorvana fanns vara en viktig faktor till att testa på att redigera första gången, och Wikipedias koppling till en mans-dominerad programmeringskultur diskuteras som en faktor till den låga andelen kvinnor. Wikipedias policies och riktlinjer och dess sammankoppling med skillnader i män och kvinnors kommunikationsstilar på internet diskuteras även.

Acknowledgements

First of all, I would like to thank all respondents to the questionnaire – thanks for contributing!

Secondly, thanks to everyone at Wikimedia Sverige for all of your time, your invaluable feedback and insights: Jan Ainali, André Costa, Axel Pettersson, John Andersson and especially my supervisor Sara Mörtsell.

Thanks to Bo Helgeson, Professor Emeritus at Blekinge Tekniska Högskola for great comments and constructive criticism.

Thanks also to all of my friends and fellow students who supported me and provided opinions and thoughts.

Special thanks to Julia Adams, Yale and Hannah Brueckner, NYU Abu Dhabi for taking time out of your busy schedule to discuss the topic with me.

Björn Helgeson

Table of Contents

Introduction

This section contains information about the degree project and an introductory background to the topic, as well as the purpose and research questions of the thesis.

Background

Wikipedia is the world's sixth most popular web site, with over 500 million visitors every month. All content of the encyclopedia is generated by unpaid volunteers, and anyone can at any time write a new article or change an existing one. Its motto has since long been: *"The free encyclopedia that anyone can edit"*. And the readers of the articles are indeed quite evenly distributed in age, gender and education levels.(Statistiska Centralbyrån 2013). However, when looking at the people writing the articles (the so called Wikipedians), studies have shown that only 1 out of 10 editors is female (Wikimedia Foundation 2011). This is a problem as Wikipedia is the most popular encyclopedia, both in Sweden and globally, and the Wikimedia Foundation strives for equality and diversity. Another more recent issue is that the quality of Wikipedia is being questioned because of the lack of diversity among its editors. In a debated article in The New York Times the author argues there's a severe imbalance between which topics are actually being covered in the encyclopedia. The author wonders why the article about friendship bracelets is so much shorter than the one on toy soldiers, and why the length of the article on a certain video game character is manifold the length of any prominent Mexican authoress'. Examples like these offer only anecdotal comparisons and are not evidence that the user gender gap affects the content; however they do voice an important concern that the content of Wikipedia does not satisfy the readers' demands. And there are also scientific investigations that indicate there is more content responding to primarily male interests (Lam et al. 2011).

This is what is commonly referred to as the Wikipedia "gender gap" – the large proportion of men in the community, and the notion that the lack of women affects the content of the encyclopedia in terms of what is covered and from which perspectives. In connection to an investigation done in 2011, the Wikimedia Foundation set up a goal to increase the portion of female writers to 25%. Despite initiatives like education, software changes and female-only workshops, the gap is not closing quickly enough. In August 2014 Jimmy Wales, cofounder of the website, admitted this goal would not be reached in time and that Wikipedia "completely failed" to address the gender issue (Hepker 2014).

In other words, the gender gap on Wikipedia is a problem from two perspectives. Firstly, it's an equality problem. The editors of Wikipedia hold a lot of power in deciding what content goes in the world's biggest encyclopedia, and from which perspective topics are portrayed. Secondly, it is a quality problem. For an encyclopedia that strives to be non-biased to have so much of its content produced by such a homogenous group is a growing concern, and risks hurting the credibility of Wikipedia as a reliable encyclopedia.

Task giver/Commissioner

Wikimedia Sweden is a non-profit organization whose main purpose is to spread knowledge. Wikimedia Sweden is the local chapter that supports Wikimedia Foundation – the foundation

that runs a number of knowledge based services and communities, including Wikipedia. Wikimedia Sweden is funded partly by the American Wikimedia Foundation, partly by donations and partly by governmental stipends and grants such as Vinnova.

Supervisor at Wikimedia Sweden:
Sara Mörtsell *sara.mortsell@wikimedia.se*
+46(0)73-383 26 70

Purpose & Goal

The purpose of the report is to firstly examine the current status of the gender gap of Swedish Wikipedia, both in terms of user-base and its effect on content. Since no prior research on the topic exists on the Swedish site, the first step is simply to attain an accurate measure of the current state. I intend to look both at users, but also at the content directly to see differences in user numbers depending on gender, and in contributions made. Secondly, the intention is to explore the reasons behind the gender gap. Previous research suggests that female survival rate in the editing community is significantly lower than their male peers' (Lam et al. 2011). However I hypothesize that this does not account for the entire gender-gap on Wikipedia. I therefore want to approach the topic with respect to two different aspects of how the gender gap has arisen. The first is why women stay a shorter period of time in the community; the second is why fewer women than men even try editing articles in the first place. With a deeper understanding of these mechanisms, suggestions can be made on which areas are the most fruitful to focus on in order to address the gender gap, to make Wikipedia a more diverse community and improve its content and coverage.

Research Questions

Q1: How big is the gender gap on Swedish Wikipedia?
Q2: Why aren't women contributing more to Wikipedia?

Delimitations

This project is primarily focused on investigating the Swedish language Wikipedia. However since it grew from the English language one, and the projects share many aspects of open collaboration; the history, conditions and some statistics from the U.S. will be used as well.

Literature review

In the following section a brief overview of Wikipedia and its history is given. The gender gap on Wikipedia will also be discussed as part of a greater problem in the field of gender and technology studies. The purpose of this section is not to exhaustively answer the research questions, but to provide a factual background from which discussion of the results from the other sections can be made. The sources are mainly found from searches in scholarly search engines like KTH Primo and Google Scholar with keywords like "gender", "Wikipedia", "women", "editors" and these works' references. For a full list of sources included in the literature review and summaries of each article, report or book, see http://wikipediaequalitythesis.blogspot.se/

Women in computing

A common notion on the state of women in computing or programming today can be expressed as: "*It's always been a male dominated field, but nowadays things are improving*". This segment contains a brief history along with the current state of women in computing, indicating the situation is closer to the opposite.

Many of the first computing pioneers were actually women. At the end of the 20th century, many of the people working on the first mechanical computational machines were mathematicians and other science academics – a not so uncommon vocation for women at the time – albeit most women holding academic degrees focused on teaching (Gürer 2002). For example, the mathematician and writer Ada Lovelace (1815-1852) is often credited as being the first programmer in history as she in the notes on her husband's mechanical analytical machine created the first computer algorithm (Fuegi & Francis 2003). In the same article she also pointed out the first formula-to-algorithm error, making her the first 'debugger' in history as well.

Incidentally, the term 'debugging' was coined by another important woman in early computing – Grace Hopper (1906-1992) – after she discovered an error caused by an actual moth inside the computer. Hopper is mentioned as one of the most important people in computer history by the magazine Business Insider (2011). Hopper was a United States Navy admiral and active in a time where the scene for women in computing would come to change a lot.

During World War II, the first electronic computers were in the making. When most of the men in the U.S. where off waging war, many women worked with engineering and science focused on aiding the war effort. One such field was ballistics computation, where hundreds of women would work double and triple shifts in computational halls, calculating artillery fire tables for the military. Therefore, when the world's first electronic reprogrammable computer ENIAC was created by the United States Army's Ballistic Research Laboratory in 1946 – all of the six programming experts hired to program the computer were women. (Gumbrecht 2011). However these women received little recognition for this work, a feat both common and part of a bigger systemic downplay of women's role in computer history, argues Light (1999), Gührer (2002) and (Ensmenger 2012).

FIGURE 1: (Unidentified U.S. Army Photographer 1946). *PICTURE OF ENIAC – THE WORLD'S FIRST ELECTRONIC COMPUTER. THE PICTURE IS FROM THE US ARMY AND HAS BEEN AFTER-TREATED WITH A DARKENING FILTER, OBSCURING OUT SOME DETAILS, SUCH AS THE WOMEN WORKING IN THE BACKGROUND.*

In 1984, when the founders and many of the would-be first community members of Wikipedia were in their adolescence, 37% of all Computer Science degrees awarded were to women. And despite the proportion of women in higher education growing over all since then, the share of women in Computer Sciences diminished rapidly over the years, standing at 26% in 1998, plateauing out at around 17-18% in 2012 (see Figure 2. In other words, the proportion of females graduating in computing decreased by 46% over the last 25 years.

In 2001, when Wikipedia was founded, around 29% of the people working in the computer field in the U.S. were female. Ten years later, in 2011, this figure has dropped to 26.6%, and if you look at the proportion of women working specifically as programmers, this figure is 23.0% (U.S. Census Bureau 2013).

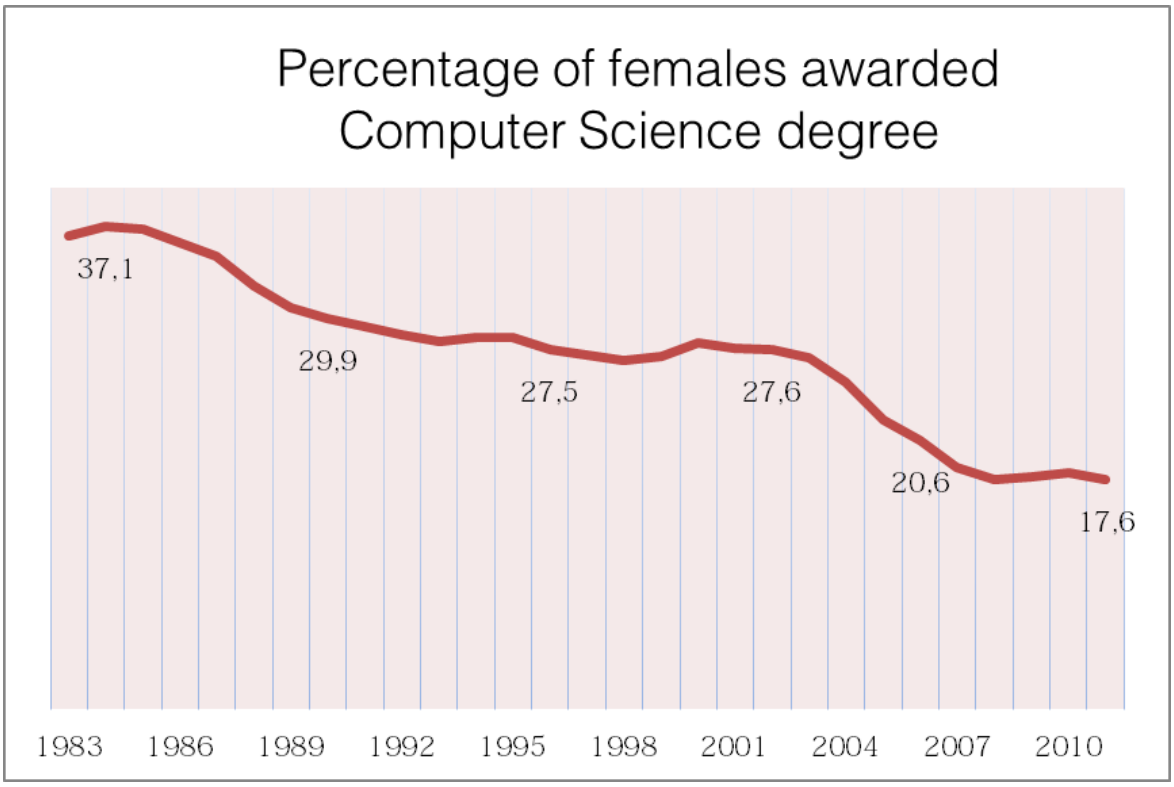


FIGURE 2: FIGURES ON AWARDED COMPUTER SCIENCE DEGREES IN THE UNITED STATES FROM NATIONAL CENTER FOR EDUCATIONAL STATISTICS

In other parts of academia, the gender gap is decreasing. Sweden has seen a lot of improvement since the dawn of the 21st century. According to statistics from the Swedish Higher Education Authority, two thirds of all bachelors degrees and 60-point Master's were awarded to women during 2012. However, the further up the academic hierarchy you get, the lesser proportion of women generally. If you look at 120-credit Master's Degrees, the ratio has dropped to 42% women. The percentage of PhD's awarded in 2013 were close to 50%, however the total amount of professors is still only around 24%, with litte improvement the last 10 years (Swedish Higher Education Authority 2014). Research has shown that both getting articles accepted for publishing by peer-reviews (Wold & Wennerås 2001) and grant-seeking (Bornmann et al. 2007) is more difficult for women. The figures in the states are similar; women earn less, are less likely to receive tenure, and appear in lesser proportions the higher up the academic ladder you go (West & Curtis 2006). Gender bias in academia is relevant to the thesis because Wikipedia's policies stipulate that all the content in the articles should be based on academic published material.

Women and computer mediated communication (CMC)

Sue Gardner, at the time executive director of the Wikimedia Foundation, acknowledged the factor of gender differences in communication styles for Wikipedia's user disparity (Gardner 2011). In research done by analyzing chat room and mailing list activities, some important differences in communication styles were mapped by Herring in 1993 and its findings updated and supported 20 years later (Kapidzic & Herring 2011), (Guiller & Durndell 2007). In short, men

show a higher propensity to state own opinions or beliefs as facts, tell others that they are wrong, and insist their own statements are correct, even if not. Women on the other hand more often express assertions in an attenuated manner, apologize, ask questions or describe their personal orientation. The following table in connection to the Wikipedia's NPOV policy will be considered further in the discussion section.

Women's Language	**Men's Language**
Attenuated assertions	Strong assertions
Apologies	Self-promotion
Explicit justifications	Presuppositions
Questions	Rhetorical questions
Personal orientation	Authoritative orientation
Supports others	Challenges others
	Humor/sarcasm

FIGURE 3: GENDER AND COMPUTER MEDIATED COMMUNICATION STYLES (HERRING 1993)

To illustrate communication styles with a real life example from the Talk/Discussion page of "Blade Runner", under the headline "Rape/Seduces":(Wikipedia article talk page 2013)

One user argues: "*In the film there is no evidence that Deckard rapes Rachel. She does consent. It could be argued that there was some duress initially, but there is no evidence of rape.*" This user exhibits strong assertiveness and stating speculation or opinions as facts.

A responding user argues: "*I'm not part of the 'rape cabal' being bandied around below, but I do think the scene is a classic rape. Deckard shoves her, holds her down, and then orders her to kiss him. Not once does she look like she's enjoying it.*" This user exhibits attenuated assertion and describes personal orientation.

Other factors

When looking at the current situation of women as computer users, there are a few more aspects related to Wikipedia editing. First off – women in the world have less access to computers and internet. In developing countries on average, 25% fewer women than men have access to internet (Intel Corporation 2013). This difference in Sweden is much lower – only around 3% across the population, except if you look at the oldest cohort. Among swedes at the age of 75-85, 66% of men have internet access, whereas this figure for women is 39% (Statistiska Centralbyrån 2013).

When looking at the current situation of women as computer users, there are a few more aspects to Wikipedia editing. First off – women in the world have less access to computers and internet. In developing countries on average, 25% fewer women than men have access to internet (Intel Corporation 2013). This difference in Sweden is much lower – only around 3%

across the population, except if you look at the oldest cohort. Among Swedes at the age of 75-85, 66% of men have internet access, whereas this figure for women is 39% (Statistiska Centralbyrån 2013).

But even though generally having have less leisure time, women spend more time in general than men on other online internet communities. In fact, already in 2012 women spent around 39% more time on social media platforms than men (Nielsen 2012). In another 2012 investigation, the demography of 24 internet communities and social media platforms (Wikipedia not included) were measured through Google Ad Planner – showing an average gender distribution of 51.25% female users, with high female ratios on social media sites like Pinterest (79%), Facebook (60%) and Twitter (60%), but much lower in computer or programming related communities like Github (30%) and Stack Overflow (24%) (Nielsen 2012). Note that the internet community with lowest ratio of females (technology news site Slashdot with 13%) still had a higher female user ratio than Wikipedia at the time.

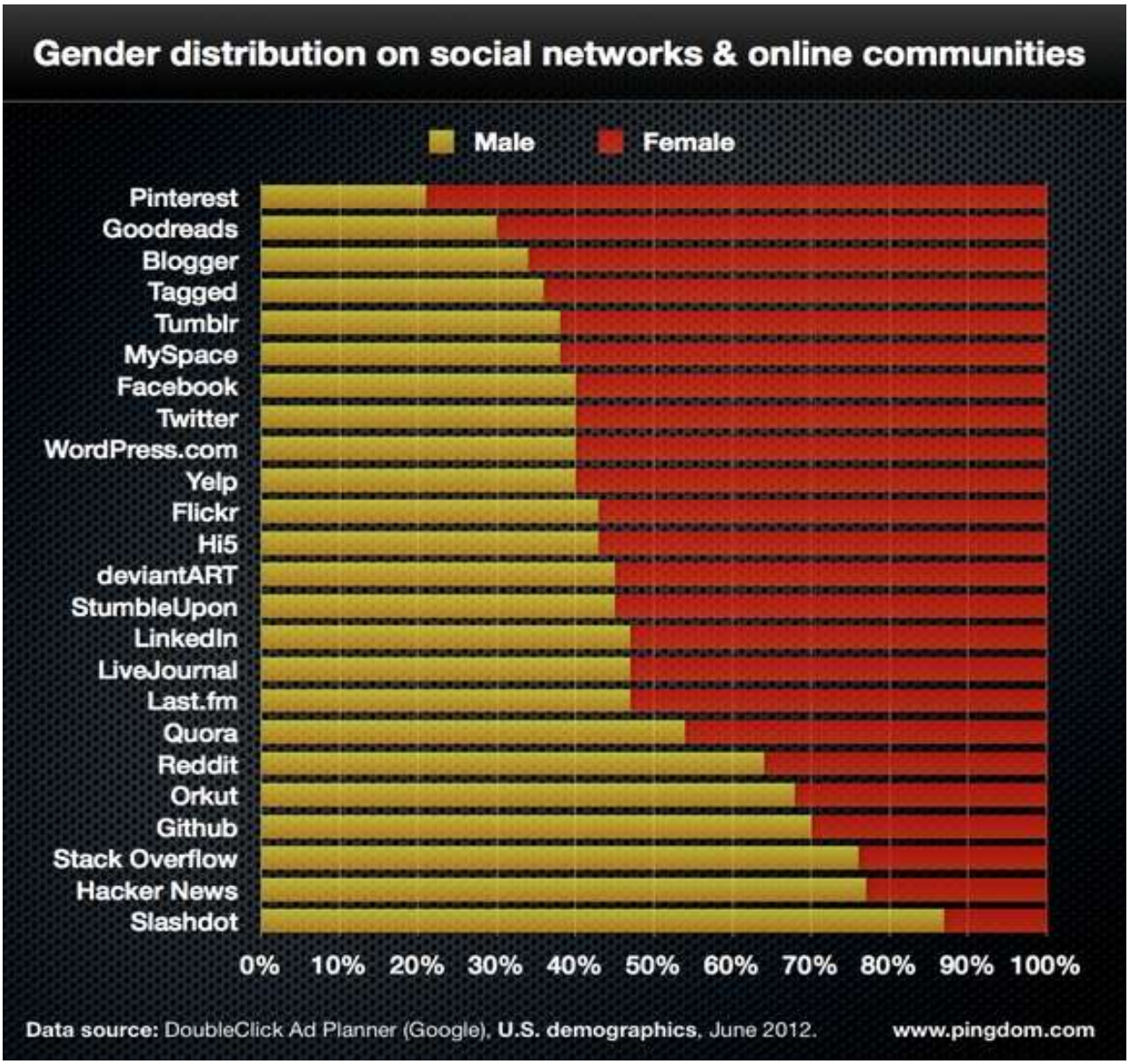


FIGURE 4: USER ACCOUNT GENDER RATIOS ON INTERNET COMMUNITIES (Pingdom Tech Blog 2012).

Swedish internet usage demographics are similar to the U.S. ones. 56% of Facebook users and 61% of Instagram users in Sweden are female (Stiftelsen för internetinfrastruktur 2014). According to Statistics Sweden (the Swedish governmental agency for official statistics), out of the people who reported that they use internet for "*posting messages to chat sites, blogs or used instant messaging*", 56% were women and 44% men. When the same sample answered "*Have posted messages to chat rooms, newsgroups or an online discussion"*, 49% were women and 51% men. (Statistiska Centralbyrån 2013)

There is little available research on the demographics of the authors of modern printed encyclopedias. For a historical perspective; in research done on the 11th edition of *Encyclopædia Britannica* (first printed in 1911), among 1500 contributors, 34 (2%) were women (Thomas 1992).

How did Wikipedia start out?

Wikipedia was launched in January 2001 by founders Jimmy Wales and Larry Sanger. It was originally intended as a complimentary resource for the Nupedia project – also a free online encyclopedia in which volunteers could contribute. However the main difference was that on Nupedia, content went through several stages of expert peer review before being approved and posted on the website.

The early users and editors of Wikipedia and Nupedia were closely connected to the Free/Libre/Open-Source Software (FLOSS) movement (Reagle 2010). Nupedia was merged with GNUpedia and Wikipedia adopted the GNU Free Documentation Licence in 2001. Jimmy Wales himself stated that "*Nupedia was absolutely inspired by the free software movement*" (Reagle 2010).

This FLOSS movement, at the time of Wikipedias formation, consisted of only 1.1% percent women (Ghosh et al. 2002). In a survey done in 2006, 1541 members of the FLOSS community were asked wether or not they had observed discriminatory behaviour against women. 78% of the men answered "No", but 75% of women said "Yes" (Krieger & Leach 2006).

In and around 2001, Wikipedia's core principles were laid out, among them the "Neutral-point-of-view" policy that constitues one of the few fundamental rules that cannot be superseded by other policies or by editors consensus. The importance of this principle will be considered further in the discussion section.

The Wikipedia project grew rapidly over the early years after 2001. In the first years there was a massive rise in the number of new editors. However, the creation of the WikiProject *Countering Systemic Bias* (Wikipedia, 2004) in 2004 can be seen as a growing concern that this influx was too homogenous a group. Five years later, by 2006, English language Wikipedia boasted its 1 millionth article. That same year, the community saw the birth of *WikiChix* - the first mailing list hosted by the Wikimedia Foundation that was not open to everyone.

"*WikiChix is a wiki and mailing list for female wiki editors to discuss issues of gender bias in wikis, to promote wikis to potential female editors, and for general discussion of wikis in a friendly female-only environment..." "...the list was organised to avoid a specific problem— women feeling uncomfortable posting to this male-dominated list where explicitly sexist statements (even if they weren't meant seriously) are left unchallenged by a large number of people.*" (Reagle 2010)

In 2010, a report was published by the United Nations University (UNU-MERIT), containing the results of the first large-scale global survey done on Wikipedia editors (Wikimedia Foundation 2011). The demography section revealed the proportion of female editors to be less than 13%. This prompted further investigation and in 2011 the foundation admitted the problem and Sue

Gardner, the executive director at the time, set a goal to increase the number of female editors to 25% by 2015 (Coen 2011).

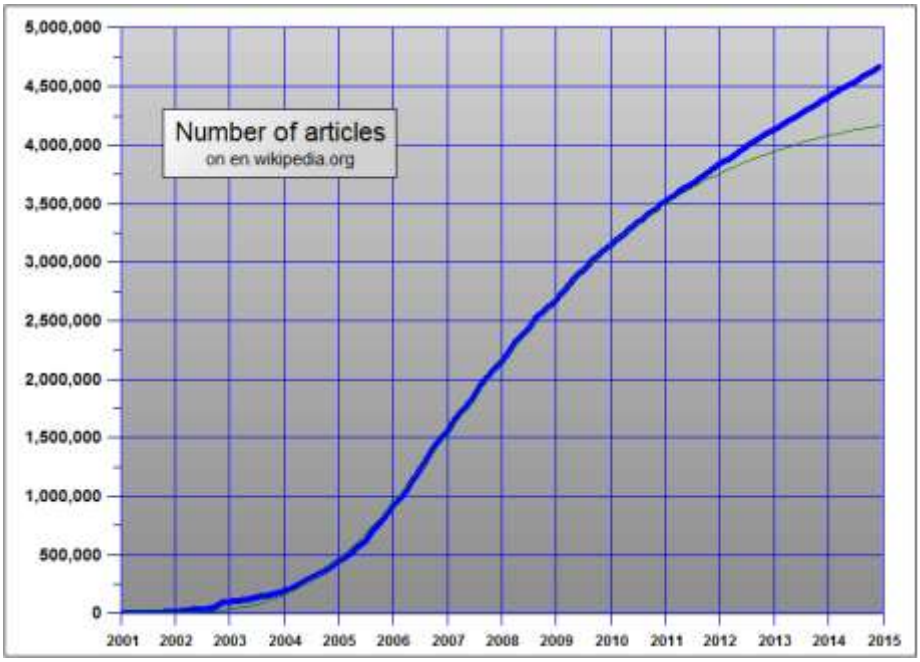


FIGURE 5: NUMBER OF ARTICLES ON WIKIPEDIA BY YEAR (WIKIPEDIA 2015).

Understanding Wikipedia

In order to understand the mechanics behind who edits and who doesn't, a brief overview will be given on how editing Wikipedia works. Wikipedia is but one of the Wiki projects run by the American-based Wikimedia Foundation (WMF). Some examples of Wikipedia's sister projects also supported by WMF include but are not limited to:

- WikiData – a linked database for structured data
- Wikimedia Commons – a repository of media files: images, sounds and videos
- Wiktionary - a lexical dictionary.
- Wikivoyage – a travel guide

The Wikimedia Foundation is a non-profit organization with the goal *"to collect and develop the world's knowledge and to make it available to everyone for free".* All the WMF-supported wiki projects are free to use and rely on crowd sourced contribution. Wikipedia is by far the largest of the Wikimedia projects, and the encyclopedia is built upon the edits of hundreds of thousands of users voluntarily contributing with information. The project is as of January 2015 available in 288 different languages, where each language has a different article content and user base. WMF doesn't have any editorial power over article content; this is all handled by the community – the voluntary users who are active in discussions and editing.

Anyone can at any time edit an existing article or create a new one – all that is required is to click the edit button in the top right corner of any article. You can then edit the contents of the article freely, changing parts of, or even the entire article. By saving your revisions, the new article version is published directly to the main page. Since the start of Wikipedia, all editing is

done with the wikimarkup language. The markup language syntax is fairly easy to learn, however it can seem intimidating to users unfamiliar to programming or markup syntax. For new users, learning this wikimarkup was the only way to edit anything on Wikipedia up until July 2013, when the simpler VisualEditor was implemented fully. VisualEditor is a more visually-based text editor extension for Wikipedia that is similar to other rich-text editors.

```
{{pp-blp|expiry=30 January 2015|small=yes}}
{{Feminism sidebar}}
{{redirect|Gendergap|other uses|Gender gap}}
'''Gender bias on Wikipedia''', otherwise known as the '''gendergap''', is the finding that between 84 and
editor surveys [http://wikimediafoundation.org/w/index.php?title=File%3AEditor_Survey_Report_-_April_2011.
[http://meta.wikimedia.org/wiki/Research:Wikipedia_Editors_Survey_November_2011 November 2011] (April - Oc
major [[criticism of Wikipedia|criticisms of Wikipedia]]. There have been attempts to narrow the gender ga
interview the [[Wikimedia Foundation]]'s plans for "doubling down" on the issue of gender bias at Wikipedi
[http://www.bbc.com/news/business-28701772 Wikipedia 'completely failed' to fix gender imbalance], [[BBC]]

== Research findings ==
'[File:Wikipedia editors are predominantly male EN.svg|thumb|300px|Wikipedia editors are predominantly mal
```

FIGURE 4: EXAMPLE OF THE WIKIMARKUP LANGUAGE, FROM THE WIKIPEDIA ARTICLE "GENDER BIAS ON WIKIPEDIA".

Many concerns about the open nature of editing are about the possibility of vandalism and that false information can be added too easily. This is indeed entirely possible and in fact happens on numerous occasions each day. However the community of serious editors is vigilant to new changes and if these don't meet the established standards and policy guidelines, chances are the edits will be reverted. Wikipedia's fundamental principles are summarized in the five "pillars":

- Wikipedia is an encyclopedia
- Wikipedia is written from a neutral point of view
- Wikipedia is free content that anyone can use, edit, and distribute:
- Editors should treat each other with respect and civility
- Wikipedia has no firm rules

Out of these, a set of policies and guidelines has been created by the community, such as sets of more detailed content and conduct policies. One example of the content policy rules on English language Wikipedia is the notability policy that governs whether or not a topic merits its own article. For example in the "people" category, the basic notability requirement is *"People are presumed notable if they have received significant coverage in multiple published secondary sources which are reliable, intellectually independent of each other, and independent of the subject."*. In addition to this there is a more detailed description of general criteria; plus specific criteria for people with different professions, as well as some exclusionary criteria.

Despite these rather extensive sets of guidelines, principles and policies there is no lack of discussions or conflicts involved in the process of editing. Each article page has a "Talk" page where editors can discuss changes to its associated article, and many entries have a manifold longer talk page than the actual article, especially for controversial topics. Most discussions on Wikipedia occur in the way the policies and guidelines dictate; they occur in a civilized manner

where proposed changes are argued for and against until, ideally, consensus is achieved. However some conflicts get so infected the involved users are unable to solve the issues themselves. Some conflicts result in so called edit wars, where users repeatedly override each other's contributions instead of reaching consensus through discussion.

Although everyone can read and access the content in the encyclopedia, there are a few user groups with special rights to perform certain actions. *Unregistered users* who have not logged in to an account have their IP addresses logged instead of usernames when making changes. They may still edit most encyclopedia pages, except those protected for certain reasons (usually controversial topics). *Logged in users* has access to these features, however newly registered users go through a short trial period. All users can partake in the discussion (and in the English language Wikipedia - voting), on which users should be granted administrative rights. *Administrators* are users with special user rights to rollback edits, delete articles, apply protection to pages or block other users. As of January 2015, there are 67 administrators on the Swedish language site and 1,362 on the English one. On the English Wikipedia there is also a higher instance called the Arbitration Committee, which consists of a panel of editors to act in editor disputes which the community otherwise is unable to resolve.

Theoretically, administrators have no more influence or power than other users when it comes to article content and they should never use their status to gain an advantage in a dispute they are involved in. However, it's hard to imagine a community in where the opinions expressed by experienced and well-known users are equal to those signed by anonymous IP addresses. Wikipedia founder Jimmy Wales stated in an interview to the New York Time in 2007:

"Greatest misconception about Wikipedia: We aren't democratic. Our readers edit the entries, but we're actually quite snobby. The core community appreciates when someone is knowledgeable, and thinks some people are idiots and shouldn't be writing." (Lewine 2007)

Another example of the politics behind editing Wikipedia is a quote from discussions regarding a mailing list, published in the book Good Faith Collaboration (Reagle 2010),:
"*There are many private, semi-private and secret lists in which wikimedians make decisions with each other without ever telling anyone or explaining. Openness has gone overboard a very long time ago. Most things you read on the public lists have been discussed privately long before an outsider found out about them*." (Walter van Kalken 2006)

There are many cases of report of administrator abuse, and in fact, Wikipedia users and administrators have been accused so many times of being secret members of different cabals, there is a humorous/satirical Wikipedia entry "List of Cabals", with more than a hundred different alleged cabals (Wikipedia 2015).

In a very recent event, Wikipedia and the Arbitration Committee was accused of censoring feminist editors who contributed to the Gamergate controversy. From an article by the Guardian:

"The editors, who were all actively attempting to prevent the article from being rewritten with a pro-Gamergate slant, were sanctioned by "arbcom" in its preliminary decision... …The byzantine internal processes of Wikipedia are incomprehensible for many, but they serve to shape the content on the site, the seventh biggest on the internet. Its reportedly unpleasant internal culture

and unwelcoming atmosphere for new editors has long been blamed for an overwhelmingly masculine make-up – just one in ten editors are thought to be female – which in turn contributes to which topics get featured on the site." (Hern 2015)

Although the Wikimedia Foundation usually doesn't comment on activity and decisions made by the community, they issued a blog post to address this critique shortly after. A quote from the post reads:

Let me close by reiterating what the Arbitration Committee's decision is not. It is not a statement on who is right or wrong regarding the Gamergate controversy article. It is not a referendum on whether Wikipedia supports or rejects feminists. The Committee's mandate is to uphold a civil, constructive atmosphere that furthers Wikipedia's mission. (Beaudette 2015)

It is a common notion that Wikipedia is one and the same encyclopedia. This is partially true, however each language version manifests in having its own editing community and cultural conditions, affecting both climate and topics. Each language has its own range of users, who discuss different topics and make different changes to their language version. For example, on Swedish language Wikipedia, there was a panel created in 2006 with the same function as ArbCom. However it was shut down in 2007, possibly for the reason that there weren't enough severe conflicts to mediate in, in order to justify having some of the administrators also performing these duties.

Key takeaways

- Even though women on the internet has increased rapidly during the last 10 years and comprise a majority on most social media sites, they are today far less represented in areas connected to computing. The proportion of women graduating with computer science degrees has decreased steadily since 1984.
- The early community of Wikipedia was closely connected to a computing culture and the FLOSS movement. It was these individuals who created the policies, values and culture that Wikipedia grew from.
- The default way to edit Wikipedia has up to 2013 been a markup syntax language.
- There are key differences in how men and women communicate on the internet, where men tend to express themselves in a more assertive manner.
- Even though there's an extensive set of rules and policies, there is no shortage of heated discussions and politics involved in editing.

Research Method

This section contains a description of the methods apart from the literature that was used to investigate the research questions. This was done by means of:
a) A database investigation of the public databases pertaining to users and editing habits.
b) A web survey distributed through the Swedish language Wikipedia aiming towards both readers and editors of the Swedish site.

Database queries

See Appendix B for a table with examples of the database queries that will be discussed further in the result section. All queries were run either through the Wikimedia Tool Labs or the publicly available web-version of the SQL-search tool Quarry (at quarry.wmflabs.org). All of the results and figures are taken from queries conducted in 2015-01-30.

Validity & Reliability (Database Queries)

The option on the user account settings page that is here equated to gender is actually formulated as: "*How do you prefer to be described?*
- I prefer not to say
- She edits wiki pages
- He edits wiki pages"

Which gender pronoun a user prefers is not necessarily the exact same as the user's gender identity, which could constitute an error. For example, a male-to-female transsexual might still prefer to be called "He", even though identifying as a woman. However such conflicting cases are estimated to be so rare they will not affect the outcome of the investigation. Also, one might assume a bias among these users with known gender settings to be those more active in the community, since the more time you spend on the website, the more likely you are to customize your settings page.

In the queries edits are used as a measure of activity for users. You could also look at number of created articles, the absolute number of bytes of each edit or calculate time spent by different categories of edits. However looking at absolute number of bytes will attribute high activity in editors focusing on cleaning up vandalism, since articles sometimes get deleted entirely by vandals, and then restored by patrolling editors, measuring high on the activity scale while taking very little time to accomplish. Looking at number of created articles is not as good a measurement either, since the majority of articles are not created in one edit by one person, but out of the collaborative edits of many editors. The chosen approach was to look at edits, since it is simple and of sufficient quantity to make accurate comparisons with. One could look at edits of different categories. For example, edits can be selected by the user as "minor", which usually means spelling corrections or adding a link to another page or such. However if looking at only minor or major edits, users with different activity preferences are measured differently, since users that do mostly patrolling (looking for vandalism, spelling mistakes, et cetera), would receive a boost or a shrinkage in their activity measures.

Questionnaire

The questionnaire was created using the survey building tool Qualtrics, with a license from the Wikimedia Foundation. It was distributed through the Swedish language Wikipedia site (sv.wikipedia.org). A link to the survey was placed in a banner that was shown to all users (both logged in and non-logged in). The banner was shown on the entire site including the front page and all articles, from 2014-12-20 to 2014-12-28, eight days.

FIGURE 6: BANNER WITH BUTTON LINKING TO SURVEY, SHOWN ON SWEDISH WIKIPEDIA WITH SUBPAGES

The banner was not shown on mobile devices, as such banners are deemed too intrusive for mobile applications. There was also a button to dismiss the banner for logged in users. Swedish language Wikipedia had about 156 million page views during December 2014, of which roughly 1/3 was through mobile devices (Wikipedia 2015). According to Wolfram Alpha's page rank tool, Wikipedia has about 700 000 unique visitors every day and 1.5 million unique visitors every week.

The questionnaire is divided into three sections. The first and last is given to all participants, while the questions of the second part differ depending on if you are currently editing, have never edited, or have edited in the past. Some of the demography questions were phrased the same as in other studies (for easy comparison), namely the Wikipedia Editor's Study 2011 and the Statistics Sweden computer use study 2013.

A precondition for distributing the questionnaire was that no questions were mandatory; all questions could be left unanswered. When a user left a question unanswered on a page, before moving on to the next page they were prompted with a warning box saying: "*There are one or many unanswered questions on this page, would you like to proceed?*" However the results of this is that in questions that include the option "No opinion", there are two ways to choose this option, either by checking the box marked "No opinion", or by leaving the question unanswered, since this too should be an acceptable way of not expressing your opinion on that particular matter.

See Appendix A for the full list of questions in the questionnaire. Noted that these are English translations of the original questions in Swedish. Also see Figure 7 for a screenshot from how the survey design and layout looked.

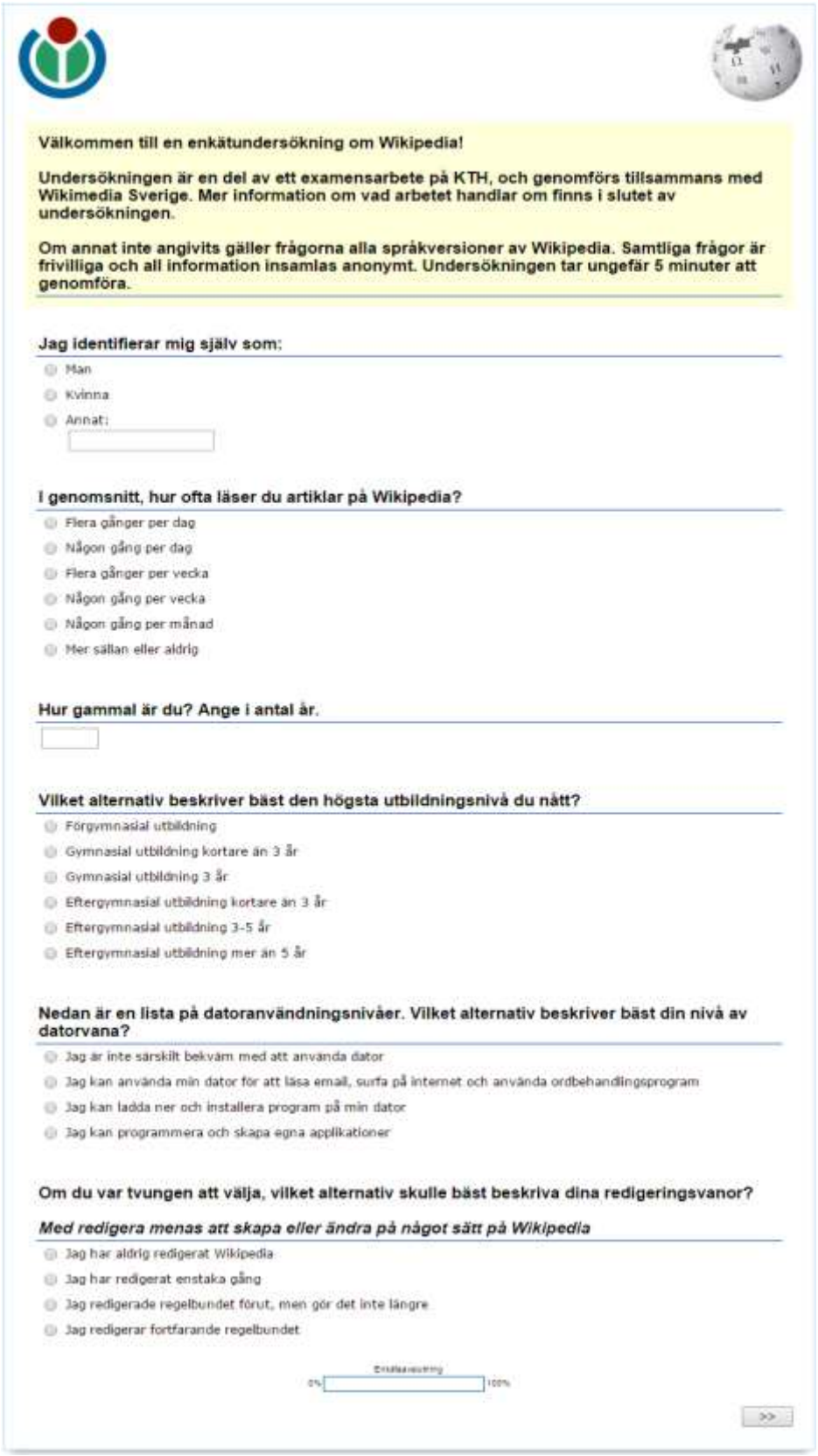


FIGURE 7: SCREENSHOT FROM THE SURVEY, PART 1 - DEMOGRAPHY

Validity & Reliability (Questionnaire)

The most prevailing inaccuracy the questionnaire will suffer from is the sampling method. Self-selection will always yield less reliable results than a probability sample with good response rates (Yeager et al. 2011). However today there is evidence that self-selection web surveys with large enough sample sizes and proper stratification can produce good representativeness (Martinsson & Andreasson 2014), (Weiss et al. 2011), (Hill & Shaw 2013) . This can also be illustrated by the success of organizations using self-selection web panels like YouGov, Ipsos and CINT. From an article in Swedish magazine Medievärlden, a representative from YouGov argues that today in countries with high internet penetration, the traditionally acclaimed telephone sampling method is not as reliable today, because more people have internet than before, and less people have home phones than before. (Björkman 2013).

Self-selection samples has also been shown to yield less reduced social desirability bias and reduced survey satisficing effects (Weiss et al. 2011). Another study supports the reliability of internet surveys when weighted by offline demographic data (Börsch-Supan et al. 2004).

Lastly for the topic of self-selected internet surveys – a very important distinction to make is that a major part of the objection to internet surveys is that they exclude non-internet users from the sample, making it less representative. However in this case the population (readers of Swedish Wikipedia) all have internet access by definition. And another difference is that if you for instance ask visitors of a newspaper how they are going to vote in an upcoming election, the responses are also going to be biased towards the political direction of the newspaper. However if you ask the newspaper about the design of their website, you would not expect such a bias to affect the outcome.

However what can be hypothesized about the response bias, is that the respondents will tend to be the more active readers and users. This because the higher the immersion, the more likely a visitor to the website will be to spend some minutes stating their opinion about the website. Also, because the banner is showed per page-view, the more times the visitors visited the page, the more likely they are to click the banner.

The survey was designed to be as concise as possible to avoid fatigue bias. In test-runs it took around 5 minutes to complete. Questions in the third part were put variedly in a reversed-key fashion to lessen acquiescence bias. No information about the topic of the thesis was included in the banner or introductory text, to avoid oversampling from respondents engaged in the topic.

Last, but not least, the survey response will be weighted with data gathered from a secondary source with probability sampling and that is not internet-based (See Figure 8). The two factors to be weighted by is age and gender. A minimum weighting-group size of 30 will be required.

58. Har sökt på wikipedia eller andra wiki-webbplatser på nätet
Andel och antal personer i åldern 16–85, avser första kvartalet 2013
58. Have used the internet to consult wikis (e.g. Wikipedia)
Percentage and number of persons aged 16–85, first quarter 2013

Ålder Utbildningsnivå Sysselsättning Familjetyp Hushållsinkomst Födelseland	Andel i procent						Absoluta tal
	Totalt	Ki±	Män	Ki±	Kvinnor	Ki±	Totalt
16–24 år	89	5	87	7	91	6	1 002 700
25–34 år	82	6	81	8	82	8	991 700
35–44 år	73	6	70	9	76	8	912 500
45–54 år	66	6	67	9	65	8	841 400
55–64 år	51	6	57	9	46	9	593 200
65–74 år	35	6	38	8	32	8	359 200
75–85 år	13	5	21	8	8	5	79 500
Förgymnasial utbildning	46	5	47	7	45	7	794 900
Gymnasial utbildning	62	3	64	5	59	5	2 347 500
Eftergymnasial utbildning	79	4	85	5	75	5	1 625 800
Studerande	89	5	84	10	92	6	859 500
Anställda/Egna företagare	72	3	72	4	73	4	3 168 000
Arbetssökande	67	13	.	.	.	.	178 700
Pensionärer och andra	28	4	34	6	24	5	574 000
Hushåll utan barn	56	3	59	4	53	4	2 957 100
Hushåll med barn	77	4	76	6	79	6	1 823 100
Hushåll med en vuxen	56	5	63	7	50	6	979 900
Hushåll med två vuxna	64	3	64	4	64	4	3 800 300
Under 25 000 kr	49	4	52	6	46	5	1 162 900
25 000 kr eller högre	72	3	72	4	72	4	3 016 100
Uppgift saknas	57	7	57	11	57	9	601 200
Inrikes födda	65	2	67	3	63	3	4 082 200
Utrikes födda	51	7	53	9	50	9	698 000
Samtliga 16–74 år	67	2	67	3	66	3	4 700 700
Samtliga 16–85 år	63	2	64	3	61	3	4 780 200

FIGURE 8: TABLE TO REPRESENT POPULATION FOR WEIGHTING RESPONSES (Statistiska Centralbyrån 2013)

Results

In this section, the findings from the database queries along with the results of the questionnaire are detailed. Only the findings relevant for this study are presented, to see more results or the entire survey raw data, go to:
https://se.wikimedia.org/wiki/Fil:Unders%C3%B6kning_Wikipedia.zip

Database Queries

The intended measurement in most cases of these queries is the activity of different user groups. For this reason, edits made in all namespaces, article content, talk pages, policy pages et cetera, are included in the measurement."

As of January 2015 there are some 28 500 000 revisions that together make up the content of Swedish language Wikipedia, among it the 3.2 million articles that constitute what is called the main namespace (the “real” content of Wikipedia – the articles in the encyclopedia). There are 415 000 registered users on the Swedish site, however only 129 000 users has made any kind of edit to the Swedish language Wikipedia. The reason there are so many unproductive accounts is because accounts are created automatically when switching from other Wiki projects, for example the English (or other) language Wikipedia.

In addition, only 91% of the revisions are made by logged in registered users. If you are not logged in or never created a user account, your revision is stored anonymously and consequently is not connected to any user, only the IP address of the editor is logged.

Also very important to note is that around 25% of all the 28 million revisions done are made by users flagged as “bots”. These are computer scripts given authority by the community to perform automated editing too repetitive or time-consuming to do manually. Examples of bot tasks can include: creating new articles with information from another database, changing all instances of linking to a certain article where naming conventions have been updated or reverting vandalism.

What’s also important to note is that only a fraction of the users has chosen an option in the gender setting on their user setting’s page. This makes it hard to estimate the gender proportions of the community, as only 5% of the users have chosen to change this setting from the default neutral value. So out of the 130 000 users there are 4900 men and 1100 women (see Figure 9).

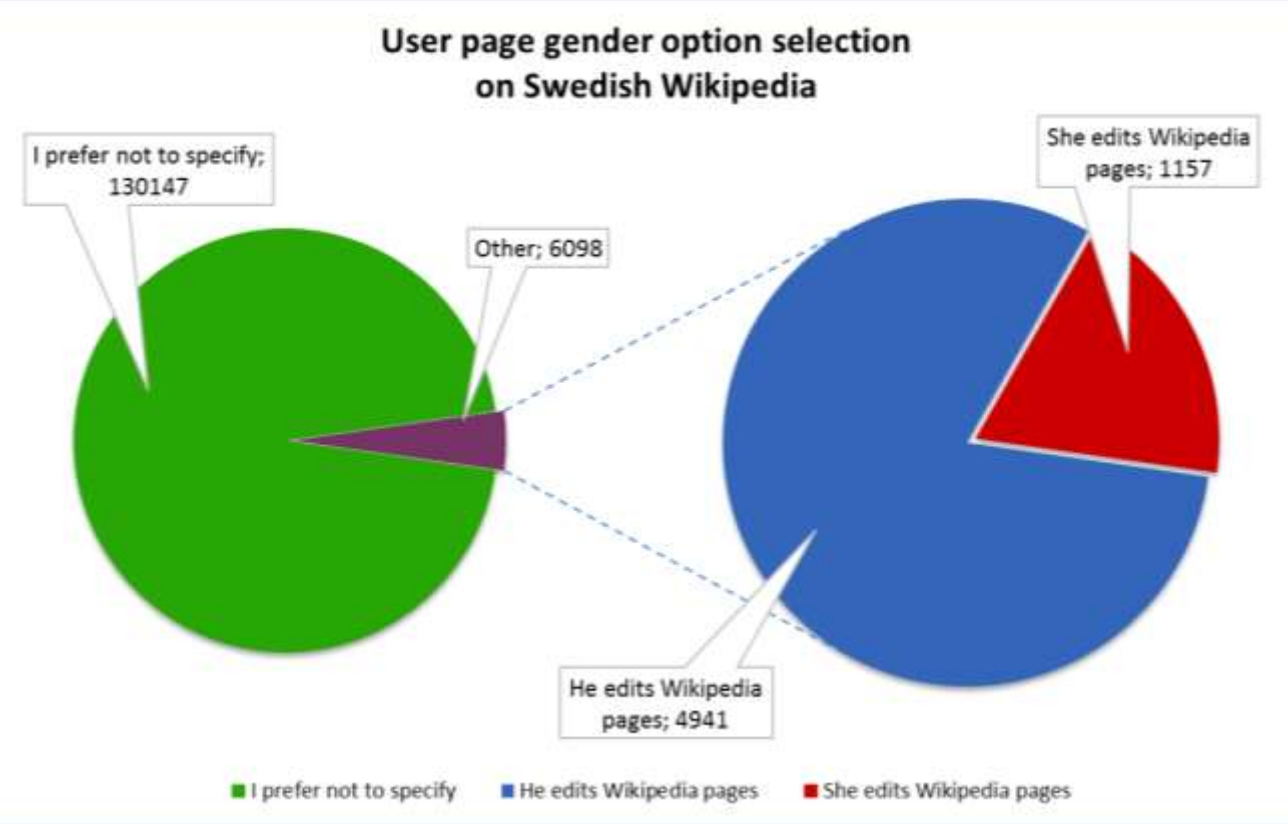
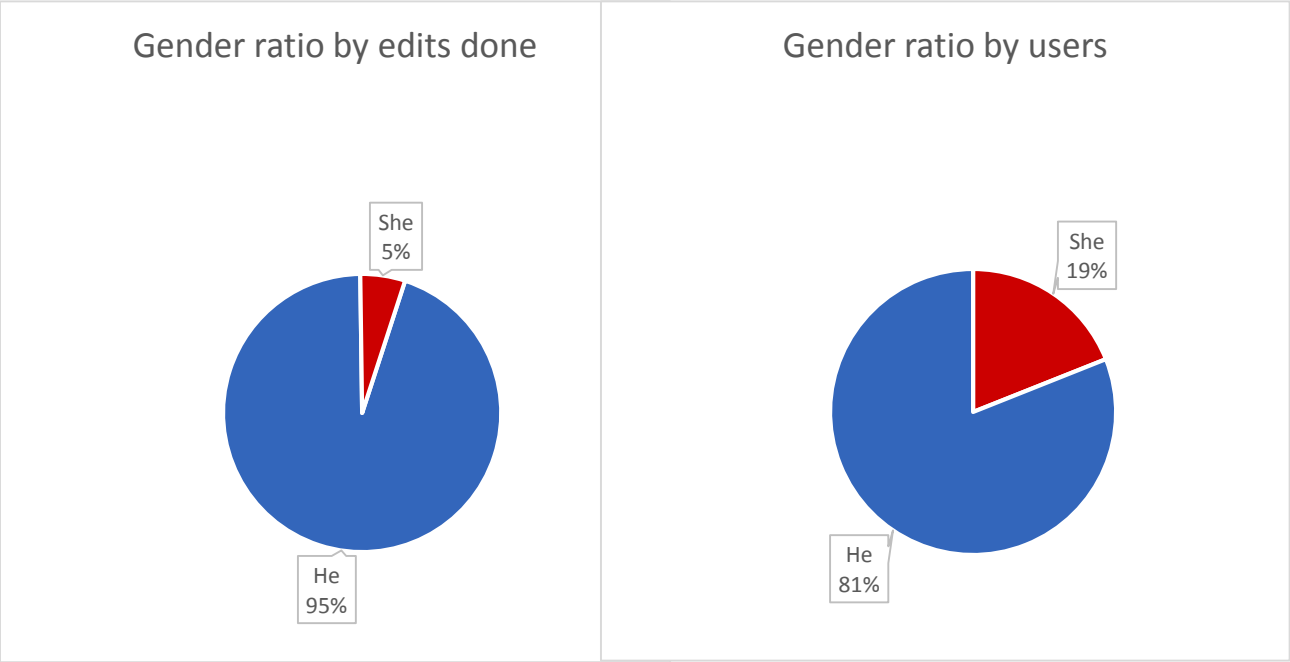


FIGURE 9: USERS ON SWEDISH LANGUAGE WIKIPEDIA

If these accounts with self-reported gender accurately represented the total population this would mean a 19,0% female-to-male ratio. Among these users, there is a clear trend in editing activity. Men average more than double the amount of edits compared to women (341 edits per man to 154 edits per woman).

These 6000 men and women together have 3.4 million edits on their record. The fact that bot accounts can have a gender setting as well complicates the database queries. There are 1.76 million edits done by users with the gender option chosen and who are not bots. Out of these "handmade" edits, only 5% are made by women (See **Error! Reference source not found.**).



Questionnaire

FIGURE 10: GENDER RATIOS BY EDITS DONE AND BY USER ACCOUNTS

In total there were 2770 completed surveys. 13 respondents who preferred not to answer the question about editing habits, were omitted. There were 61 respondents who chose “Other” in the gender question. Even though it is an interesting and important group to study, which has been neglected in much of scientific history, unfortunately this sample group is too small for comparisons. Also, many answers in the free text questions from this group were unserious, hence I have chosen to omit these cases. The respondents who marked the option “I have never edited Wikipedia” are termed *readers*, where those who chose “*I still edit regularly*” are addressed as editors in the following pages.

Which alternative best describes your editing history?

By editing means to create or change in any way on Wikipedia

		Options				
		I have never edited Wikipedia	I've edited a few times	I edited regularly before, but I am not editing anymore	I still edit regularly	Total
	Man	1064	714	122	186	2086
	Woman	403	140	16	29	588
Total		1467	854	138	215	2674

FIGURE 11: GENDER AND EDITING HABITS AMONG RESPONDENTS

Gender & Age Group Distribution of readers

	Age Group							
	16-24	25-34	35-44	45-54	55-64	65-74	75-85	Total
Man	227	80	90	155	204	211	91	1058
Woman	114	41	34	56	57	64	34	400
Total	341	121	124	211	261	275	125	1458

FIGURE 12: GENDER AND AGE GROUP DISTRIBUTION OR RESPONDENTS WHO MARKED "I HAVE NEVER EDITED WIKIPEDIA". EACH CELL HAS A SAMPLE SIZE > 30. DATA PRESENTED IS BEFORE WEIGHTING.

In comparison to the data from Statistics Sweden, there was a skewness towards the oldest age groups, the youngest age groups and male respondents. The skewness is in line with the hypothesis that the respondents are more active visitors than average.

Gender & Age Group Distribution of active editors

	Age Group							
	16-24	25-34	35-44	45-54	55-64	65-74	75-85	Total
Man	30	20	36	40	28	24	5	183
Woman	4	3	5	7	7	1	0	27
Total	34	23	41	47	35	25	5	210

FIGURE 13: GENDER & AGE GROUP DISTRIBUTION OF USERS THAT REPORTED THEY ARE STILL ACTIVE EDITORS

Among the respondents who answered “I still edit actively” on the editing habits question, there were 13% women (See Figure 13). The number of active editors that answered this questionnaire is very large compared to the total number of active editors (around 800 active editors per month).

Unknown-gendered users appears non-divergent

In the questionnaire the respondents who had made edits to Wikipedia (N=1250) were asked about how they treated the gender pronoun question in the user account settings page. This was done in order to gain a better understanding of the large amounts of accounts with unspecified gender settings. The result was that 14.9% of the women respondents of the questionnaire reported having their gender specified in the settings box (N=302) and 16.1% of the men (N=938). As seen in Figure 14, there were also a very low portion who had explicitly chosen the opposite gender (only 3 cases out of 1240) and a low percentage of people who actively chose to not specify gender (4.8%).

In the user settings page on Wikipedia there's an alternative called "How would you like to be addressed?"

	Which option did you choose?					
	I checked "She edits Wikipedia pages"	I checked "He edits Wikipedia pages"	I chose actively to check "I prefer not to specify"	Never saw the setting/left is as it was	Never created a user account	Total
Man	1	112	38	503	284	938
Woman	21	2	22	142	115	302
Total	22	114	60	645	399	1240

FIGURE 14: GENDER OPTION SELECTION CHOICES BY SELF-REPORTED SURVEY GENDER

Since the percentages are similar for both genders in this questionnaire, this supports the hypothesis that the unknown users on Wikipedia are distributed similarly to the users with known gender settings.

Computer skills important factor for trying out editing

Wikipedia's connection to programming culture is explored further by asking the respondents to share their perceived computer skill levels (since no tests of their actual computer skills were made). As seen in Figure 16, computer skills are similar across all groups having tried out editing at least a few times, and lower among readers who never tried editing. The amount of readers who can program & create their own applications are 17.0%, whereas this figure is around the double for all other groups.

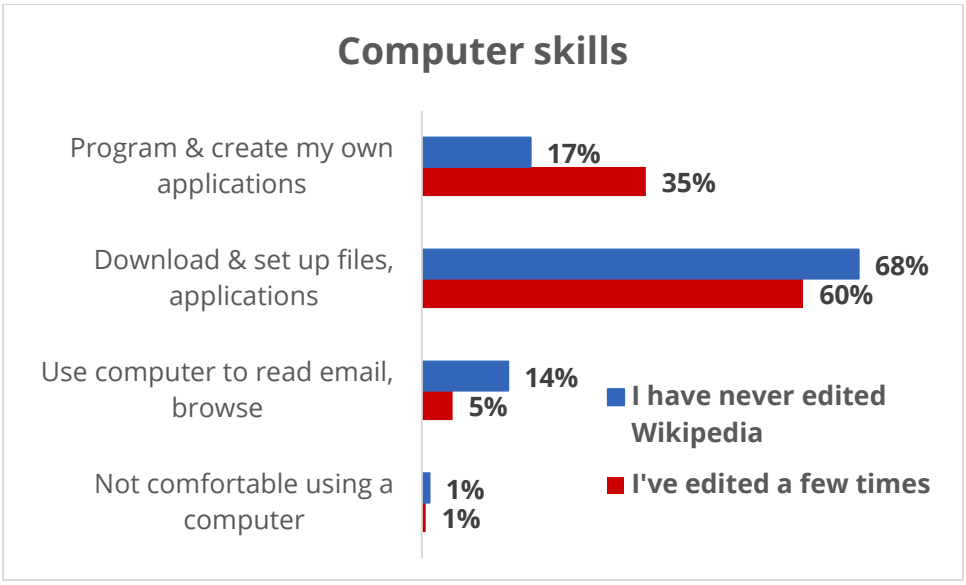


FIGURE 15: COMPUTER SKILL COMPARISON BETWEEN RESPONDENTS WHO HAVE TRIED EDITING A FEW TIMES, AND THOSE WHO HAVEN'T.

Below is a list of computer proficiency levels. Which alternative best describes your level of proficiency?

	Not comfortable using a computer	Use computer to read email, browse	Download & set up files, applications	Program & create my own applications
I have never edited Wikipedia	18	198	1000	249
	1,2%	13,5%	68,3%	17,0%
I've edited a few times	6	45	564	331
	0,6%	4,8%	59,6%	35,0%
I edited regularly before, but I am not editing anymore	1	10	89	57
	0,6%	6,4%	56,7%	36,3%
I still edit regularly	2	14	136	74
	0,9%	6,2%	60,2%	32,7%
Total	27	267	1789	711
	1,0%	9,6%	64,0%	25,4%

FIGURE 16: PERCEIVED COMPUTER SKILLS, BY EDITING HABITS.

This is an important figure, because when looking at these differences by gender (see), men are almost three times as likely to know programming. 25% of men contra 9% of women readers consider themselves able to program & create their own applications. In essence: those who tried editing Wikipedia have higher computer skills than those who never tried, and generally men's perceived computer skills are much higher. These results are in line with previous findings from the literature studies about women in computing declining.

Below is a list of computer proficiency levels. Which alternative best describes your level of proficiency? (readers only)

		Not comfortable using a computer	Use computer to read email, browse	Download & set up files, applications	Program & create my own applications	
Men		10	70	488	186	754
		1,3%	9,3%	64,7%	24,7%	100,0%
Women		8	128	512	63	711
		1,1%	18,0%	72,0%	8,9%	100,0%

FIGURE 17: PERCEIVED COMPUTER SKILLS AMONG READERS, BY GENDER

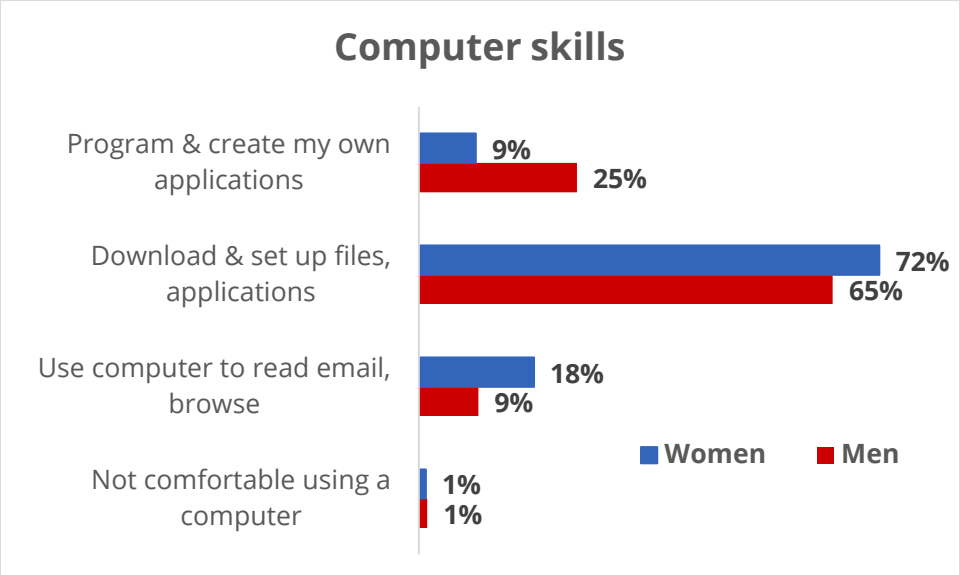


FIGURE 18: CHART OF PERCEIVED COMPUTER SKILLS AMONG READERS, BY GENDER.

Reasons for not contributing similar across gender groups, barring one question

The respondents who had never edited Wikipedia were asked to what extent a list of reasons had contributed to them never editing. A Chi-Square test on the 8 different questions presented showed there was a statistically significant difference between genders in only one of the presented options ($p<0.000$, $N=1119$). Women selected that "*I'm not competent enough to edit*" *Contributes strongly* almost twice as often as men (See Figure 18).

To what extent do these reasons contribute to you never having edited Wikipedia?

		-I'm not competent enough to edit			
		Does not contribute	Contributes somewhat	Contributes strongly	Total
	Man	281	196	91	568
	Woman	237	145	169	551
Total		518	341	260	1119

FIGURE 19: FREQUENCY TABLE FROM QUESTIONNAIRE QUESTION - OPTION "I'M NOT COMPETENT ENOUGH TO EDIT"

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	34,515[a]	2	,000
Likelihood Ratio	34,904	2	,000
Linear-by-Linear Association	19,392	1	,000
N of Valid Cases	1119		
Mann-Whitney U test "-I'm not competent enough to edit"			
Mann-Whitney U	121496,000		
Wilcoxon W	273572,000		
Z	-4,010		
Asymp. Sig. (2-tailed)			,000

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 128,03.

FIGURE 20: STATISTICAL TESTS FOR QUESTIONNAIRE QUESTION OPTION "-I'M NOT COMPETENT ENOUGH TO EDIT", CHI-SQUARE & MAN-WHITNEY U

Conceptions about Wikipedia among readers overall positive

The third and final part of the questionnaire consisted of questions about how readers – people outside of the community – regard the community itself, and a series of Likert scale items were presented with 5 options ranging from "Strongly Disagree", "Disagree", "No opinion", "Agree" and "Strongly Agree". As visualized in Figure 22, the readers appear to have an overall positive attitude towards Wikipedia and its community. The option "No opinion" is truncated in the figure, but account for the missing percentages for each question.

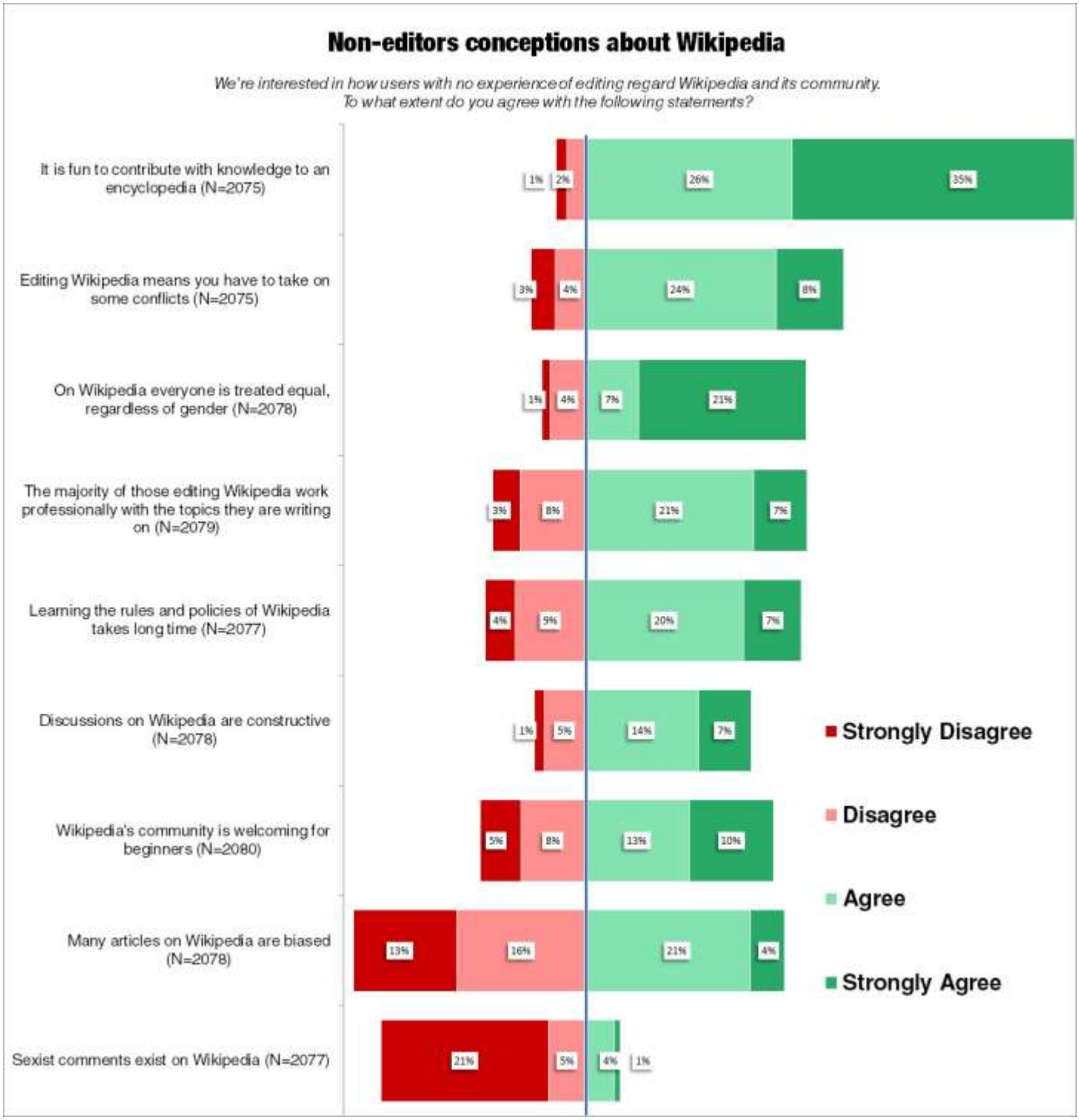


FIGURE 21: READERS ANSWERS FROM QUESTIONNAIRE ABOUT THEIR CONCEPTIONS ABOUT WIKIPEDIA. THE OPTION "NO OPINION" IS TRUNCATED IN THIS GRAPH, BUT ACCOUNT FOR THE REST OF THE ANSWERS.

The statement the respondents agreed most strongly with, was that "*It is fun to contribute with knowledge to an encyclopedia*" (N=2075). 32% of the respondents also agreed or agreed strongly with the statement that "*Editing Wikipedia means you have to take on some conflicts*", indicating that there is an awareness among readers about discussions and disagreements behind the editing of the articles. A conception among readers seems to be that the majority of those editing Wikipedia work professionally with the topics they are writing on, as 28% percent agreeing (agreed or agreed strongly), with 11% disagreeing. This was also the question where editors' and readers' answers differed the most. Among editors this figure was 32% disagreeing and 35% agreeing.

Both editors and readers disagree with the statement that sexist comments exist on Wikipedia. But probably readers are far less likely to read the discussion pages and only read the articles. However readers also agreed largely with the statement that "*On Wikipedia everyone is treated equal, regardless of gender*", indicating they don't have any particular preconceived notion of sexism in the community.

The question with the greatest dispersal was "*Many articles on Wikipedia are biased"*, with 29% disagreeing and 25% agreeing. Furthermore, the readers favor slightly the agreeing side on questions "*Wikipedia's community is welcoming for beginners*" and "*Learning the rules and policies of Wikipedia*".

We are interested in what people's perceptions are of the Wikipedia community - in other words what your preconceptions are about the users that edit Wikipedia today. To what extent do you agree with the following statements?

Mann-Whitney U Test - Ranks

	Gender	N	Mean Rank	Sum of Ranks
-Wikipedia's community is welcoming for beginners	Man	607	603,29	366198,50
	Woman	555	557,67	309504,50
	Total	1162		
-On Wikipedia everyone is treated equal, regardless of gender	Man	606	603,46	365698,50
	Woman	555	556,47	308842,50
	Total	1161		
-Discussions on Wikipedia are constructive	Man	606	583,99	353898,50
	Woman	555	577,73	320642,50
	Total	1161		
-Editing Wikipedia means you have to take on some conflicts	Man	605	615,45	372349,50
	Woman	554	541,28	299870,50
	Total	1159		
-The majority of those editing Wikipedia work professionally with the topics they are writing on	Man	605	569,80	344728,50
	Woman	556	593,19	329812,50
	Total	1161		
-Learning the rules and policies of Wikipedia takes long time	Man	605	578,38	349917,00
	Woman	556	583,86	324624,00
	Total	1161		
-Many articles on Wikipedia are biased	Man	605	597,25	361333,50
	Woman	555	562,25	312046,50
	Total	1160		
-It is fun to contribute with your own knowledge to an encyclopedia	Man	603	591,58	356725,50
	Woman	554	565,30	313177,50
	Total	1157		
-Sexist comments exist on Wikipedia	Man	605	554,93	335731,00
	Woman	555	608,38	337649,00
	Total	1160		

FIGURE 23: STATISTICAL TESTS ON OPTIONS TO QUESTION REGARDING PEOPLE'S PERCEPTIONS OF WIKIPEDIA

Mann-Whitney U Test[a]

	Welcoming for beginners	Equal, regardless of gender	Discussions constructive	Editing means conflicts	Editors work professionally on editing topics	Learning policies takes time	Articles biased	It is fun to contribute	Sexist comments exist
Mann-Whitney U	155215	154553	166353	146136	161414	166602	157757	159443	152416
Wilcoxon W	309505	308843	320643	299871	344729	349917	312047	313178	335731
Z	-2,673	-2,736	-,396	-4,256	-1,290	-,320	-1,862	-1,409	-3,249
Asymp. Sig. (2-tailed)	,008	,006	,692	,000	,197	,749	,063	,159	,001

a. Grouping Variable: Woman/Man

FIGURE 24: RANK SUMS FOR MANN-WHITNEY U TESTS.

The questionnaire results reveal significant gender differences in conceptions about Wikipedia on 4 accounts:

- Women are less likely to think Wikipedia is welcoming to beginners than men ($p<0,008$)
- Women are less likely to think everyone gets treated equally, regardless of gender ($p<0,006$)
- Women are less likely to state that editing means taking on conflicts ($p<0,001$)
- Women are more likely to acknowledge the existence of sexist comments ($p<0,001$)

On the question "*Which alternative best describes your knowledge about how editing Wikipedia works?"* almost half of the readers (46%) answered "*I know nothing about how editing Wikipedia works*". Only 7% of the respondents stated "*I know very well how editing Wikipedia works*".

Discussion

This section contains discussions of the results and the literature study. A portion of this section contains discussions on the rules and policies of Wikipedia coupled with other research from the literature study. The results from the database queries and questionnaire are then interpreted and evaluated in a larger context with regards to literature and research.

Gender gap on Swedish Wikipedia exists

First of all, the findings support that there is a widespread gender gap on the Swedish language Wikipedia as well. By using three data sources – the database user account settings, the questionnaire respondents distribution and the questionnaire user account settings question - a more accurate estimation of the severity of the gender gap is made. Firstly, when looking at users with known gender settings in the database there are 19% women on Swedish language Wikipedia (N=6098). However looking solely at users' distribution alone is not sufficient to make an accurate conclusion, since they constitute such a small portion of all editors. Secondly, coupling this figure with data from the questionnaire, asking how users handled their user settings on the account page, allows for the sampling of users with unknown account gender settings. The results indicate that the gender distribution among users with unknown gender settings does not deviate from the distribution of known-gendered users that was measured in the database queries to be 19%. And thirdly, when looking at the gender distribution of current or past editors among respondents to this survey, the proportion of women is around 13% (N=353). These results indicate the editor gender gap indeed prevalent in the Swedish language Wikipedia as well, and in the same proportion as on the English language site.

And this despite many considering Sweden being a country with good gender equality. Swedish Wikipedia cannot attribute its gender gap to lesser internet access, less free time and lower education levels to the same extent as other language Wikipedias. And as in other parts of the world, the internet gender gap has decreased substantially over the last decade, to the point of women being a majority in most popular social media communities in Sweden – for example Facebook, Instagram, Twitter, Pinterest and Tumblr.

This makes it extra peculiar that the gender gap on Wikipedia is so large and seem to be closing so slowly, and prevalent even in countries like Sweden. My view is that this is connected to the computing and programming culture that was discussed in earlier chapters. As seen in the result section, computer skills are an important factor for trying out editing, perhaps to get past the initially complicated-looking markup language. This could be an important factor to the gender gap, since women's perceived computer skills are generally lower.

So although we might have expected a rise in number of women on Wikipedia during the last 10 years because of increased internet access and social media immersion, this is counteracted by a larger declining trend in women in computing and programming.

Wikipedia's policies and guidelines

The "No original research policy" and the scientific community

The "No original research" policy - one of Wikipedia's core content principles - postulates that in the articles the editor should not strive to share their own opinion or provide their own statements, but rather the content should reflect the state of the current scientific community's published works. As mentioned earlier in the literature review, gender proportions and attitudes are already skewed in the scientific community and this impacts who gets funding to conduct their research and who gets to publish their work. This results in a fundamental fallacy; even if the "No original research" policy were to be followed flawlessly without other biases, the content and discussions on Wikipedia would still be based on a male dominated scientific community and its values on different scientific disciplines, methods and researchers.

NPOV policy and communication styles

A second problem with Wikipedia's core policies is that certain communication styles are favored. The purpose of Wikipedia's "Neutral Point of View" policy is to discourage biases and editors' own opinions and values to end up in the article, to ensure there is nothing but facts in the encyclopedia. And this works very well at a basic level – an example from the NPOV article on Wikipedia: "*an article should not state that 'genocide is an evil action', but it may state that 'genocide has been described by John X as the epitome of human evil.'"*. But of course most discussions are not of this kind of textbook nature. There are many examples of complex and problematic situations where different sides of a dispute have a hard time resolving conflicts just by quoting published material. Some examples of disputes related to gender in the editing community in the past has been:

- The aforementioned Gamergate controversy. For example, should the article state that the topic is about ethics in journalism or sexism in the gaming community? Is it biased to call it a controversy?
- Should a debated scene in the movie Blade Runner be described as a rape scene or a consentual sex scene?
- Is it androcentric to have the category "American female novelists" and the corresponding category for males called "American novelists"?
- Should the article about the U.S. army soldier who leaked documents to WikiLeaks and later underwent a sex change be titled "Bradley Manning" or "Chelsea Manning"?

What these examples illustrate is that following the fact-centric rules and policies will only get you so far, the community will eventually have to make decisions where the lines between facts and opinions are blurred. In the above examples it is hard, if not impossible, to favor one decision over another based on fact rather than opinion. I argue that there are many situations like these where decisions have to be made more based on opinion than on fact. And since the rules and policies advocate expressing only facts, it is the editors who argue most factual-like in these opinion-based issues that are favored. In other words, those who disguise their own opinions as facts the best will be able point to the policies and say "your argument is original research, mine is a fact".

To go back to the example mentioned in the literature review chapter with the Blade Runner scene; two users are discussing their impressions and how to interpret a scene in a movie. They both base their statements on the exact same information. One writes assertive statements such as "*She does consent*", while the other acknowledges their own limitations in interpreting and not being able to know for sure, writing "*Not once does she look like she's enjoying it"*. The users both argue based on the same information, however users expressing opinion as fact have support in the "Neutral point of view"-policy and can dismiss the others arguments as "original research" – which actually happened later on in the discussion in this example.

To connect this back to gender and CMC – an important difference between genders in communication styles is that men more often and successfully state opinions as facts, and communicate with a more assertive style, regardless of actually being right or wrong. (Kapidzic & Herring 2011). Women more frequently ask questions, make attenuated assertions or describe their personal orientation rather than make statements. So when editing articles, discussing changes, categorizing information and choosing administrators, the policies will favor the male communication style.

And also, since the majority of editors are men, it is men who makes the distinction of which expressions are opinions and which are facts. It is not necessarily so that it is active misogyny and that editors are always aware of the sender of the message or statement. That being said, this does not mean the policies could simply be rewritten without biases or favoring certain communication styles. Having another policy that supports caution when expressing certainty or somehow punished over-assertiveness would firstly be very hard to conceive of and implement, and secondly, bring other problems and fallacies.

Gender gap and number of unknown cases

The database user setting query indicates 19% of the users of Wikipedia are women. As the number of unknown cases is so high, one cannot draw the conclusion that this figure is representative of the entire population of editors. Results from the questionnaire however, showed that there was no discernible difference between genders in the reporting of this setting. Of the questionnaire respondents who were still active users there were 13% women. The reason for this percentage to be slightly lower can be explained by an overrepresentation of active users as a part of self-selection bias. As the more active users tend to be men and more active users have a higher propensity to answer the questionnaire, this corresponds well with the figures. All of these findings are consistent with the conclusion that there is no hidden higher portion of either gender among the unknown users – the user gender gap on Swedish language Wikipedia seems to be as wide as initially perceived. And when looking directly at contributions made by known-gendered users, if they accurately represent the unknown-gendered users, the encyclopedia is made up by 95% male contributions.

Readers' conceptions

From the first part of the questionnaire, readers were asked what reasons contributed to them not editing. Out of these 8 questions, only one showed a statistically different difference across gender groups. Women were far more likely than men to choose "*Not competent enough to edit*"

as contributing to them not editing Wikipedia. This is in line with other studies on gender and risk and conflict aversion (Jianakoplos & Bernasek 1998), (Eckel & Grossman 2008)

The results from these 8 questions indicate no particular preconceived notion of sexism among readers. The questionnaire results does not find any evidence for gender differences in to what extent the readers experience the editing interface too complicated, nor in fear that their edits would be reverted. In fact, around 45% of both women and men stated that they know nothing about how editing Wikipedia works, despite the strong formulation of this alternative.

From the third part of the questionnaire; there were significant differences between gender groups in what conceptions are held about Wikipedia and its community. These differences can be summarized as: women have more concerns about the community being sexist and not welcoming, and do not expect conflict as part of editing to the same degree as men. However the differences in the Whitney Mann U test ranks were relatively small, indicating the differences in opinion between gender groups do not differ greatly. This, in combination with the very low total percentage that had negative expectations of the community, makes it difficult to say that this is an important contributing factor to women not editing Wikipedia. For example, only around 1% of the total respondents strongly agreed that sexist comments exist and that men and women are not treated equally – but among these respondents, a significantly large portion were women. However this does not mean one can completely exclude these reasons from playing a bit part in explaining the gender gap. Is possible that small differences in welcoming of different groups, combined with a higher propensity to attract same-sex friends to the community could have played a big role in the forming of the gender gap.

However, these actual results, combined with the findings from the question asked earlier chronologically (about reasons for not contributing) do not indicate that preconceived negative notions about the community contribute to the gender gap. This is an important finding, since part of the purpose of the questionnaire was to pick up any strong particular preconceptions among readers that might be inhibiting their willingness to contribute.

Conclusion

This section summarizes the findings and discussion in six key takeaway points, then proceeds to answer the research questions explicitly. Suggestions for further research is also provided.

Key takeaways:

1. **There is a significant gender gap on Swedish language Wikipedia.** The findings of this study suggest that between 13-19% of editors are women.
2. **Few who have tried editing experienced the community as unwelcoming or sexist**. However among those who did, a significantly large portion were women.
3. **Significant differences in perceived competence were found.** Women report. “I’m not competent enough” is a strong contributing factor to them not editing more than twice as often as men.
4. **Even though the proportion of women in most online communities are increasing, the percentage of women in programming has declined.** As Wikipedia at its founding was connected a programming culture, this may be a reason for the resilience of the gender gap.
5. **Even though the proportion of women in most online communities are increasing, the percentage of women in prograamming has declined.** As Wikipedia at its founding was connected a programming culture, this may be a reason for the resilience of the gender gap.
6. **Even though the proportion of women in most online communities are increasing, the percentage of women in programming has declined.** As Wikipedia at its founding was connected a programming culture, this may be a reason for the resilience of the gender gap., which is frowned upon.

Research Questions

Q1: How big is the gender gap on Swedish Wikipedia?

All findings from the questionnaire and database queries are consistent with the conclusion that there is a prevalent gender gap on Swedish Wikipedia, of the same order as in the English language one. Among user accounts with registered gender settings there are 19% women in the database. From the questionnaire, among the editors or previously active editors in the survey there were 13% women respondents. From the questions regarding gender account settings there were no indications of a different gender proportions among users with unknown gender settings. To summarize, the findings of this study indicate that the gender proportions on Swedish language Wikipedia is 13-19% women and 81-87% men.

Q2: Why aren't women contributing more to Wikipedia?

Even though some readers perceived the community as sexist, unequal or unwelcoming to beginners, the overall conceptions of Wikipedia and its climate were positive, and the findings do not support the hypothesis that preconceived negative notions are a strong factor of the gender gap. One prominent factor discussed is Wikipedia’s origin as strongly connected to a male-dominated programming culture. Perceived computer skills emerged as an important factor for trying to edit the first time, which is relevant since women’s perceived computer skills were generally lower in both readers and editors. The increase in proportions of women in other

parts on the internet combined with other gender equalizing factors can be seen as counteracted by Wikipedia's connection to programming, and it being a sector with rare declining women portions.

Suggestions for further research

I believe it would be interesting to further investigate women and computer mediated communication styles in connection to Wikipedia's gender gap. And to make further investigations of the issue through the distinction of two problems would be interesting; that women are less likely to stay in the community once immersed, and that women are less likely to enter the community in the first place. Through regression analysis one could make an estimate of how many editors would be women if the survival rates in the community would be the same.

It could also be insightful to get more data on perceived vs. actual computer skills in Sweden, and measure with greater detail the perceived and actual computer skills of Wikipedia editors and non-editors, to further explore the connection between computing and the gender gap.

References

Björkman, L., 2013. Det finns inget slumpmässigt urval längre. *Medievärlden*. Available at: http://www.medievarlden.se/diskussion/2013/05/det-finns-inget-slumpmassigt-urval-langre [Accessed January 21, 2015].

Bornmann, L., Mutz, R. & Daniel, H., 2007. Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1(3), pp.226–238. Available at: http://www.sciencedirect.com/science/article/pii/S1751157707000363 [Accessed February 16, 2015].

Börsch-Supan, A. et al., 2004. *How to make internet surveys representative: A case study of a two-step weighting procedure*, Available at: http://www.mea.mpisoc.mpg.de/uploads/user_mea_discussionpapers/loil50ozz320r55b_pd1_040330 geschuetzt.pdf [Accessed November 4, 2014].

Coen, N., 2011. Define Gender Gap? Look Up Wikipedia's Contributor List. *The New York Times*. Available at: http://www.nytimes.com/2011/01/31/business/media/31link.html?_r=0.

Eckel, C.C. & Grossman, P.J., 2008. Men, Women and Risk Aversion: Experimental Evidence. Available at: http://papers.ssrn.com/abstract=1883693 [Accessed February 16, 2015].

Ensmenger, N.L., 2012. *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*, Available at: http://books.google.se/books/about/The_Computer_Boys_Take_Over.html?id=VCcsTPQ738oC&pgis=1 [Accessed December 26, 2014].

Fuegi, J. & Francis, J., 2003. Lovelace & babbage and the creation of the 1843 "notes." *IEEE Annals of the History of Computing*, 25(4), pp.16–26. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1253887 [Accessed December 26, 2014].

Gardner, S., 2011. Nine Reasons Women Don't Edit Wikipedia (in their own words). Available at: http://suegardner.org/2011/02/19/nine-reasons-why-women-dont-edit-wikipedia-in-their-own-words/ [Accessed December 1, 2015].

Ghosh, R.A. et al., 2002. *Free/Libre and Open Source Software: Survey and Study*, Available at: http://www.flossproject.org/report/FLOSS_Final4.pdf.

Guiller, J. & Durndell, A., 2007. Students' linguistic behaviour in online discussion groups: Does gender matter? *Computers in Human Behavior*, 23(5), pp.2240–2255. Available at: http://www.sciencedirect.com/science/article/pii/S0747563206000458 [Accessed February 14, 2015].

Gumbrecht, J., 2011. Rediscovering WWII's female "computers." *CNN*. Available at: http://edition.cnn.com/2011/TECH/innovation/02/08/women.rosies.math/.

Gürer, D., 2002. Pioneering women in computer science. *ACM SIGCSE Bulletin*, 34(2), p.175.

Hepker, C., 2014. Wikipedia “completely failed” to fix gender imbalance. *BBC*. Available at: http://www.bbc.com/news/business-28701772.

Hill, B.M. & Shaw, A., 2013. The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. *PloS one*, 8(6), p.e65782. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0065782 [Accessed February 15, 2015].

Intel Corporation, 2013. *Women and the Web*, Available at: http://www.intel.com/content/dam/www/public/us/en/documents/pdf/women-and-the-web.pdf.

Jianakoplos, N.A. & Bernasek, A., 1998. ARE WOMEN MORE RISK AVERSE? *Economic Inquiry*, 36(4), pp.620–630. Available at: http://doi.wiley.com/10.1111/j.1465-7295.1998.tb01740.x [Accessed February 16, 2015].

Kapidzic, S. & Herring, S.C., 2011. Gender, Communication, and Self-Presentation in Teen Chatrooms Revisited: Have Patterns Changed? *Journal of Computer-Mediated Communication*, 17(1), pp.39–59. Available at: http://doi.wiley.com/10.1111/j.1083-6101.2011.01561.x [Accessed January 12, 2015].

Krieger, B. & Leach, J., 2006. *Free/Libre and Open Source Software: Policy Support*, Available at: http://flosspols.org/deliverables/FLOSSPOLS-D16-Gender_Integrated_Report_of_Findings.pdf.

Lam, S. (Tony) K. et al., 2011. WP:Clubhouse? An Exploration of Wikipedia’s Gender Imbalance. In *WikiSym 2011*. Mountain View, CA: ACM Digital Library.

Light, J.S., 1999. When computers were women. *Technology and Culture*, 40(3), pp.455–483.

Love, D., 2011. The 21 Most Important Names In Computing History You’ve Never Heard. *Business Insider*. Available at: http://www.businessinsider.com/history-of-computing-2011-8?op=1#ixzz3N15TObTz.

Martinsson, J. & Andreasson, M., 2014. *Olika metoder - lika resultat*, Available at: https://docs.google.com/file/d/0B_Fff1bUl_0IemZMRlBUeDB4WFE/edit [Accessed December 9, 2014].

Nielsen, 2012. *STATE OF THE MEDIA: THE SOCIAL MEDIA REPORT*, Available at: http://www.nielsen.com/us/en/insights/reports/2012/state-of-the-media-the-social-media-report-2012.html.

OECD Social Indicators, 2009. *Special Focus : Measuring Leisure in OECD Countries*, Available at: http://www.oecd.org/berlin/42675407.pdf.

Pingdom Tech Blog, 2012. *Social network demographics in 2012*, Available at: http://royal.pingdom.com/2012/08/21/report-social-network-demographics-in-2012/ [Accessed January 7, 2015].

Reagle, J.M., 2010. *Good faith collaboration*, MIT Press.

Statistiska Centralbyrån, 2013. *Privatpersoners användning av datorer och internet 2013*, Available at: http://www.scb.se/Statistik/_Publikationer/LE0108_2013A01_BR_IT01BR1401.pdf.

Stiftelsen för internetinfrastruktur, 2014. *Svenskarnas användning av sociala medier 2014*, Available at: http://www.internetstatistik.se/rapporter/svenskarnas-anvandning-av-sociala-medier-2014/.

Swedish Higher Education Authority, 2014. *Gender equality in Swedish higher education*, Available at: http://english.uka.se/statisticsfollowup/genderequality.4.614ce6ad149128a3c9e934.html [Accessed October 2, 2015].

Thomas, G., 1992. *A position to command respect*, Scarecrow Press.

U.S. Census Bureau, 2013. *Disparities in STEM Employment by Sex, Race, and Hispanic Origin*, Available at: http://www.census.gov/prod/2013pubs/acs-24.pdf.

Unidentified U.S. Army Photographer, 1946. ENIAC2. Available at: http://ftp.arl.army.mil/ftp/historic-computers/jpeg/eniac2.jpg.

Weiss, R., Krosnick, J.A. & Yeager, D.S., 2011. More comparisons of Probability and Non-Probability Sample Internet Surveys: The Dutch NOPVO Study. - WebSM. In *The American Association for Public Opinion Research (AAPOR) 66th Annual Conference*. Available at: http://www.websm.org/db/12/14387/Bibliography/More_comparisons_of_Probability_and_NonProbability_Sample_Internet_Surveys_The_Dutch_NOPVO_Study/ [Accessed January 15, 2015].

West, M.S. & Curtis, J.W., 2006. *AAUP Faculty Gender Equity Indicators*, Available at: http://www.aaup.org/reports-publications/publications/see-all/aaup-faculty-gender-equity-indicators-2006 [Accessed February 1, 2015].

Wikimedia Foundation, 2011. *Wikipedia Editors Study*, Available at: http://upload.wikimedia.org/wikipedia/commons/5/51/Editor_Survey_Report_April_2011.pdf .

Wikipedia, 2015. Page Views for Wikipedia, Mobile, Raw data. *stats.wikimedia.org*. Available at: http://stats.wikimedia.org/EN/TablesPageViewsMonthlyOriginalMobile.htm [Accessed February 13, 2015].

Wikipedia, 2004. WikiProject Countering Systemic Bias - Archived Discussion. Available at: https://en.wikipedia.org/wiki/Wikipedia_talk:WikiProject_Countering_systemic_bias/Archive_1#Page_title_move_to:_Wikipedia:WikiProject_Countering_systemic_bias [Accessed January 15, 2015].

Wikipedia article talk page, 2013. Talk:Blade Runner - Wikipedia. Available at: https://en.wikipedia.org/wiki/Talk:Blade_Runner#Rape.2FSeduces [Accessed January 9, 2015].

Wold, C. & Wennerås, A., 2001. *Nepotism and sexism in peer-review*, Psychology Press. Available at: https://www.google.se/books?hl=en&lr=&id=T_omcYIJx_8C&pgis=1 [Accessed February 16, 2015].

Yeager, D.S. et al., 2011. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, 75(4), pp.709–747. Available at: http://poq.oxfordjournals.org.focus.lib.kth.se/content/75/4/709 [Accessed January 8, 2015].

Appendices

Appendix A: Questionnaire questions

Questions Part 1: Viewed by all participants

Welcome to a survey about Wikipedia!

This questionnaire is a part of a degree project at the Royal Institute of Technology together with Wikimedia Sverige. More information about what the project is about can be found at the end of the survey. Unless specified otherwise, all questions apply to all language versions of Wikipedia. All questions are voluntary to answer, and all data is gathered anonymously.

I identify myself as:

- ❍ Man
- ❍ Woman
- ❍ Other: _____

On average, how often do you read articles on Wikipedia?

- ❍ Several times a day
- ❍ Once or twice a day
- ❍ Several times per week
- ❍ Once or twice per week
- ❍ Once or twice per month
- ❍ Less frequently/Never

How old are you?

Which alternative best describes the highest level of education you have completed?

- ❍ Primary
- ❍ High school shorter than 3 years
- ❍ High school 3 years
- ❍ College, less than 3 years
- ❍ College, 3-5 years
- ❍ College, more than 5 years

Below is a list of computer proficiency levels. Which alternative best describes your level of proficiency?

- ❍ Not comfortable using a computer
- ❍ Use computer to read email, browse
- ❍ Download & set up files, applications
- ❍ Program & create my own applications

Which alternative best describes your editing history? By editing means to create or change in any way on Wikipedia

- ❍ I have never edited Wikipedia

- ❍ I've edited a few times
- ❍ I edited regularly before, but I am not editing anymore
- ❍ I still edit regularly

Questions Part 2: Viewed only by participants who has never edited

Which alternative best describes your knowledge about how editing Wikipedia works?

- ❍ I know nothing about how editing Wikipedia works
- ❍ I roughly know how it works
- ❍ I know very well how it works

Do you know anyone currently editing Wikipedia?

- ❍ Yes
- ❍ No (as far as I know)

What, if any, are your preconceptions regarding the community that exists on Wikipedia today? (the people who edit actively

- ❍ Write here (adjectives): _____
- ❍ No opinion

To what extent do these reasons contribute to you never editing?

	Does not contribute	Contributes somewhat	Contributes strongly	No opinion
I was unaware I could edit	❍	❍	❍	❍
I don't know how to edit	❍	❍	❍	❍
I don't have enough time	❍	❍	❍	❍
I'm not competent enough to edit	❍	❍	❍	❍
Haven't discovered anything missing	❍	❍	❍	❍
The user interface to edit seems to complicated	❍	❍	❍	❍
The risk that what I write would be deleted is too large	❍	❍	❍	❍
It is too much trouble learning the rules and guidelines	❍	❍	❍	❍
I don't think I would appreciate the culture among Wikipedians	❍	❍	❍	❍

Which of these social media platforms have you shared material on, in any way, during the last three months? (for example uploaded an image, updated status or written a post on)

- ❑ Facebook
- ❑ Instagram
- ❑ Twitter
- ❑ Flashback
- ❑ Familjeliv
- ❑ Blogg
- ❑ Other Internet community

Questions Part 2: Viewed only by participants who have edited in the past, but not anymore

How long ago did you make your last edit?

- ❍ More than 2 years
- ❍ 1-2 years
- ❍ 6-12 months
- ❍ 1-6 months
- ❍ Less than 1 month

Roughly how many edits have you made in total on Wikipedia?

My edits have been done mainly in connection to an assignment (for example in school, writing workshops or similar)

- ❍ Yes
- ❍ No

How was your general experience of editing Wikipedia

- ❍ Positive
- ❍ Negative

Which reasons contributed to you stopping to edit Wikipedia? (you can mark several options)

- ❑ My edits were not accepted by other users
- ❑ Negative experience of discussions or conflicts
- ❑ Don't have enough competences to contribute further
- ❑ Changes in my life situation unrelated to Wikipedia
- ❑ Rules and policies have become too complicated
- ❑ Rather spend time on other internet communities
- ❑ Don't have enough time
- ❑ Other reasons (please specify) _____

In the user settings page on Wikipedia there's an alternative called "How would you like to be addressed?"Which option did you choose?

- ❍ I checked "She edits Wikipedia pages"
- ❍ I checked "He edits Wikipedia pages"
- ❍ I chose actively to check "I prefer not to specify"
- ❍ Never saw the setting/left is as it was
- ❍ Never created a user account

Questions Part 2: Viewed only by participants has is currently editing

Roughly how many edits have you made on Wikipedia in total?

In the user settings page on Wikipedia there's an alternative called "How would you like to be addressed?"Which option did you choose?

- ❍ I checked "She edits Wikipedia pages"
- ❍ I checked "He edits Wikipedia pages"
- ❍ I chose actively to check "I prefer not to specify"
- ❍ Never saw the setting/left is as it was
- ❍ Never created a user account

Questions Part 3: Viewed by all participants

We are interested in what people's perceptions are of the Wikipedia community - in other words what your preconceptions are about the users that edit Wikipedia today. To what extent do you agree with the following statements?

	Strongly Disagree	Disagree	No opinion	Agree	Strongly Agree
Wikipedia's community is welcoming for beginners	❍	❍	❍	❍	❍
On Wikipedia everyone is treated equal, regardless of gender	❍	❍	❍	❍	❍
Discussions on Wikipedia are constructive	❍	❍	❍	❍	❍
Editing Wikipedia means you have to take on some conflicts	❍	❍	❍	❍	❍
The majority of those editing Wikipedia work professionally with the topics they are writing on	❍	❍	❍	❍	❍
Learning the rules and policies of	❍	❍	❍	❍	❍

Wikipedia takes long time					
Many articles on Wikipedia are biased	❍	❍	❍	❍	❍
It is fun to contribute with your own knowledge to an encyclopedia	❍	❍	❍	❍	❍
Sexist comments exist on Wikipedia	❍	❍	❍	❍	❍

That was the last question! Do you have any other thoughts or comments you would like to add?

Appendix B: Database queries

Semantic Query	SQL Query	Result
DBQ1: How many of the users are women/men?	`SELECT COUNT(DISTINCT rev_user) AS WomenTotal FROM svwiki_p.revision WHERE rev_user IN (SELECT user_id FROM svwiki_p.user JOIN svwiki_p.user_properties ON user_id=up_user WHERE up_property = 'gender' AND up_value = 'female'); SELECT COUNT(DISTINCT rev_user) AS MenTotal FROM svwiki_p.revision WHERE rev_user IN (SELECT user_id FROM svwiki_p.user JOIN svwiki_p.user_properties ON user_id=up_user WHERE up_property = 'gender' AND up_value = 'male'); SELECT COUNT(DISTINCT rev_user) as Total FROM svwiki_p.revision WHERE rev_user IN (SELECT user_id FROM svwiki_p.user);`	Women: 1157 Men: 4941 Total: 130153 / 64296
DBQ2: How many edits have the women/men done in total?	`SELECT COUNT(DISTINCT rev_id) AS TotalRevisionsByMenNotBots FROM svwiki_p.revision WHERE rev_user IN ( SELECT DISTINCT up_user FROM svwiki_p.user_properties JOIN svwiki_p.user_groups ON ug_user=up_user WHERE up_property="gender" AND up_value="female"/"male" AND ug_group<>'bot' )`	Women: 72 342 Men: 1 685 553
DBQ3: How many awake* users are there today and how many edits have they contributed with in total?	`USE svwiki_p; SELECT COUNT(user_id) as Users, SUM(user_editcount) AS Edits FROM svwiki_p.user WHERE user_id IN ( SELECT DISTINCT r.rev_user FROM revision r WHERE r.rev_timestamp>20141030000000 AND r.rev_user NOT IN ( SELECT ug.ug_user FROM user_groups ug WHERE ug.ug_group='bot') )`	Users: 6774 Edits: 8 752 874

DBQ4: What is the percentages of awake to sleeping* users?	`SELECT count(DISTINCT u.user_id) as TotalUsers FROM svwiki_p.user u`
	`JOIN svwiki_p.revision r ON u.user_id=r.rev_user`
	`JOIN svwiki_p.user_properties up ON u.user_id=up.up_user`
	`WHERE u.user_id NOT IN (SELECT r.rev_user FROM svwiki_p.revision r WHERE r.rev_user IS NOT NULL AND r.rev_timestamp>20141030000000)`
	`AND up.up_property="gender" AND up.up_value="female"`
	`AND user_id NOT IN (`
	`SELECT ug.ug_user FROM svwiki_p.user_groups ug`
	`WHERE ug.ug_group='bot')`
		Women: 0,27%**
		
		Men: 1,83%

DBQ5: How many revisions were made by bots?	`SELECT count(rev_id) AS TotalRevisionsByBots FROM svwiki_p.revision`
	`WHERE rev_user IN`
	`(SELECT ug_user FROM svwiki_p.user_groups`
	`WHERE ug_group=""bot"")`
		7 167 286

FIGURE 25: DATABASE QUERIES AND RESULTS.

* = A USER IS COUNTED AS SLEEPING WHEN NO EDITS HAS BEEN MADE THE LAST 3 MONTHS.

** = TOO FEW CASES TO PROVIDE ACCURATE DATA