

Expressions of Sentiments during Code Reviews: Male vs. Female

Rajshakhar Paul*, Amiangshu Bosu*, Kazi Zakia Sultana†,

* Department of Computer Science

Wayne State University, Detroit, MI, USA

† Department of Computer Science and Engineering

Montclair State University, Montclair, NJ, USA

r.paul@wayne.edu, amiangshu.bosu@wayne.edu, sultanak@montclair.edu

Abstract—As most of the software development organizations are male-dominated, many female developers encounter negative workplace experiences and report feeling like they “do not belong”. Exposures to discriminatory expletives or negative critiques from their male colleagues may further exacerbate those feelings. The primary goal of this study is to identify the differences in expressions of sentiments between male and female developers during various software engineering tasks. On this goal, we mined the code review repositories of six popular open source projects. We used a semi-automated approach leveraging the name as well as multiple social networks to identify the gender of a developer. Using SentiSE, a customized and state-of-the-art sentiment analysis tool for the software engineering domain, we classify each communication as negative, positive, or neutral. We also compute the frequencies of sentiment words, emoticons, and expletives used by each developer.

Our results suggest that the likelihood of using sentiment words, emoticons, and expletives during code reviews varies based on the gender of a developer, as females are significantly less likely to express sentiments than males. Although female developers were more neutral to their male colleagues than to another female, male developers from three out of the six projects were not only writing more frequent negative comments but also withholding positive encouragements from their female counterparts. Our results provide empirical evidence of another factor behind the negative work place experiences encountered by the female developers that may be contributing to the diminishing number of females in the SE industry.

Index Terms—gender issues, gender bias, sentiment, code review, emotion, discrimination

I. INTRODUCTION

Until 1960’s, programming was considered women’s job and the computing area was dominated by women [5]. However, as the demand for programmers increased, male programmers sought to create their dominance through creating professional associations, through ad campaigns discouraging the hiring of women, and by adding personality tests biased against female applicants [26]. As the area of computing has thrived over the last five decades, the participation of women in computing has declined. According to the US Bureau of Labor Statistics, females accounted for about 37% of the U.S. college students who received bachelor’s degrees in Computer and Information Sciences (CIS) in 1986 [26]; twenty years later, in 2006, that number declined to only 12% [75]. The percentage of females in the software industry also declined rapidly during

the same period [60] and most of the contemporary software development organizations are male-dominated [58].

Recently, female software developers reported not only sexual harassment [23] but also various technical biases. For example, women are often assigned menial tasks [25] and code commits from females are less likely to be accepted [65] than males. Marwick *et al.* claims that many females in the current software industry often feel like they “do not belong” and therefore struggle with an “*imposter syndrome*”¹ [49]. Exposures to discriminatory expletives or negative critiques from their male colleagues may further exacerbate those feelings.

There are several anecdotal evidence suggesting demeaning attitude from males towards female developers. For example, a female developer from the Github mentioned in an interview with the Techcrunch [72],“.. *really hard time getting used to the culture, the aggressive communication on pull requests and how little the men I worked with respected and valued my opinion.*” She “*participated in the boys club upon joining,*” but later found “*her character being discussed in inappropriate places like on pull requests and issues.*” Another memo [22] written by a Google employee claimed “*Differences in distributions of traits between men and women may in part explain why we don’t have 50 percent representation of women in tech and leadership.*” However, no study has yet investigated whether or not, female developers are more frequent recipients of negative opinions than their male colleagues. We believe such an investigation can possibly determine another factor behind the low participation of female developers in the contemporary software industry. Therefore, the primary objective of this study is *to identify the differences in expressions of sentiments between male and female developers during various software engineering tasks.* However, the goal of our research is not to investigate the differences between the inherent writing styles of males and females, rather we aim to find out if they express sentiments differently during software engineering interactions and whether their opinions differ based on the gender of the recipient. If we find differences, in a followup study, we want to investigate whether the behavior

¹ A psychological phenomenon in which they feel a sense of inadequacy despite being perfectly competent.

is resulting from any social context or gender bias or from a technological standpoint.

Expressions of positive or negative opinions during various Software Engineering (SE) activities are common as prior SE research found sentiments in commit messages [31], issue tracking systems [52], code review comments [6], Stack Overflow posts [18] and mailing-lists [32]. While no study has yet explored the differences in sentiment expressions between male and female software developers, research in other domains suggest existences of significant differences based on the gender of an author. For example, females express significantly more positive sentiments on social media [66], use emoticons differently than males [73], and emotion-based features in micro-blog posts can be used to predict the gender of an author [51]. As prior research shed lights on the differences of opinions between males and females in other domains, we are motivated to investigate this question in the SE field.

Among the various software engineering interactions, such as code reviews, bug discussions, code commits, and Stack-Overflow discussions, this study focuses on code reviews. We selected code reviews as those facilitate direct communications between developers and are one of the primary interactions where developers may express sentiments [16]. For this research, we mined the code review repositories of six popular open source software (OSS) projects and identified all the developers that have committed at least five code changes for those projects. Based on the approach adopted in a recent study [65], we developed a semi-automated methodology followed by a manual validation using social networks (i.e., LinkedIn, Google Plus, Facebook, Github, and Twitter) to identify the genders of the ‘Non-casual developer’². We applied SentiSE, a customized and state-of-the-art sentiment analysis tool for the SE domain to identify the sentiment polarity of each code review comment.

In summary, the primary contributions of this study are:

- A comparison of the opinions expressed by male and female software developers during code reviews.
- A comparison of the different categories of sentiment words used by male and female developers.
- A comparison of the emoticons, expletives and swear words used by male and female developers.
- A comparison of the opinions expressed during same-gender and cross-gender interactions.
- An empirical evidence of the differences between male and female developers during SE interactions.

The remainder of the paper is organized as follows. Section II provides background about code review and sentiment analysis. Section III introduces the research questions of this study. Section IV describes our research methodology. Section V presents the results of this study. Section VI discusses the implications of the results. Section VII describes the threats to

²A developer who has submitted at least five code changes to a project under our study.

validity of our findings. Finally, Section VIII provides some directions for future work and concludes the paper.

II. BACKGROUND

This section presents a brief background on three topics relevant to this study: Gender issues in Software Engineering, Sentiment analysis, and peer code reviews.

A. Gender Issues in Software Engineering

Earlier studies in software engineering focused on the participation of women in Free/Libre/ Open Source Software (FLOSS) projects and reported extremely lower (i.e., between 2% to 5%) ratio of female contributors [30]. The reason behind the lower participation of female is attributed to the social and cultural arrangements of FLOSS projects, which actively excludes female contributors [53]. Due to several initiatives from few of the FLOSS projects (i.e., Debian, and GNOME) to attract female contributors, the number of female contributors has improved to around 11% in recent years [63]; however the number is still less than half of the industry average (i.e., 24%).

Similar to other domains [19], [34], SE researchers have also observed the positive impacts of gender diversity on software development team productivity [68], [56]. Yet, studies also found various types of discrimination against women. For example, women in computing organizations are often assigned menial tasks, while similarly male colleagues are given ‘choice’ projects [25], often do not get opportunities at management positions, and earn lower salaries compared to men [58]. Women also often receive unfair evaluation as 72% women reported sensing gender bias in how they were evaluated [35]. As a result, female developers become increasingly pessimistic about career opportunities as their tenure progress [39]. Even in FLOSS projects, where most of the participants are volunteers, women perform majority other types of contributions than coding, while men mostly contribute with code [63]. A recent study on Github reports that although women significantly have higher acceptance rate than men in general, if a woman’s gender is identifiable from her profile, her pull requests are less likely to be accepted [65].

B. Sentiment Analysis

Sentiment analysis is the computational process of determining the emotional tone behind a series of words to gain an understanding of the attitudes, opinions and emotions of a speaker or an author [47]. Sentiment analysis applications have spread to almost every possible domain, from consumer products, services, healthcare, and financial services to social events and political elections [27].

SE researchers have recently applied sentiment analysis on various SE artifacts and found developers expressing sentiments in code commits, issue reports, and project forums. Pleatea *et al.* [61] analyzed sentiments of discussions on GitHub using NLTK [15] and found more expressions of negative sentiments in security-related discussions than in other discussions. Guzman *et al.* [31] used SentiStrength to

analyze commit messages on Github and observed projects developed in Java having more negative commit messages than projects developed in other languages. They also observed higher expressions of positive sentiments among distributed teams. Garcia et al. [29] carried out a similar analysis on both the issue reports and mailing-list discussions of the Gentoo community. They observed significant correlations between the activities of authors and their expressions of emotions. Murgia et al. [52] analyzed issue reports from the Apache Foundation projects and observed developers expressing sentiments during bug discussions.

Most of the prior works focused on identifying sentiments expressed during various SE activities. Couple of recent works have evaluated the impact of those sentiments on the outcomes of different activities. Ortu et al. [55] analyzed the relation between bug resolution time with sentiment, emotions and politeness of developers and found that issues with negative sentiments are often associated with longer resolution times. Islam et al. [37] investigated emotional variations during various development activities (e.g., bug-fixing tasks) and found that emotional active developers tend to post longer commit messages.

C. Code Review

Peer code review is a software engineering practice, where a developer sends his/her code to a peer to identify possible defects before merging to the project codebase. Compared with the traditional heavy-weight inspection process, peer code review is more informal, tool-based, and used regularly in practice [10]. To make peer code reviews more efficient, teams use automated support tools such as Gerrit³, Phabricator⁴, and ReviewBoard⁵. A tool-based code review process starts when an author creates a *patchset* (i.e. all files added or modified in a single revision) along with a description of the changes and submits that information to a code review tool. To facilitate reviews, code review tools highlight the changes between two revisions in a side-by-side display. Both the reviewers and the author can insert comments pointing out issues, suggesting improvements, or clarifying the change. After the review, the author may upload a new patch-set addressing the review comments and initiate a new review iteration. This review cycle repeats until either the reviewers approve the change or the author abandons it. Code review tools capture the interactions (a.k.a. review comments) between the author and a reviewer to facilitate *post-hoc* analyses.

III. RESEARCH QUESTIONS

The primary objective of this study is *to identify the differences in expressions of sentiments between male and female developers during various software engineering tasks*. Following subsections introduce five specific research questions to investigate this objective. As this study focuses specifically on

code reviews, our research questions are aimed towards code review comments.

A. Gender Vs. Sentiment

Traditional believes in the Western culture portrait women as “the emotional sex” and consider men as emotionally inexpressive compared to women [74]. Recent research on social media posts [51], [66] also found females expressing more sentiments than males. Therefore, our first research question investigates if these findings are also applicable to professional workplace interactions such as code reviews.

RQ1: *Does the likelihood of expressing sentiments during code reviews depend on the gender of a developer?*

B. Same Gender Vs. Cross-Gender Interaction

Prior works found significant differences between same-gender and cross-gender interactions [73], [64]. In general, we would expect a developer to be less expressive, when interacting with a person from the opposite gender than with a person from the same gender. Our next research question investigates this expectation.

RQ2: *Do developers express sentiments differently during their cross-gender interactions than during their same gender interactions?*

C. Gender Vs. Sentiment Word

The study on gender and language began in the early 1970s and was arguably established by Robin Lakoff [43]. The book worked on the foundational ideas about gendered language. She outlined some specific tendencies of wording or writing styles in women’s language. Although there are many controversies and criticisms against her work, this groundbreaking research unmasked the nature of male supremacy. Another recent work found females authoring more happy / sad tweets than males, while males’ tweets expressed more surprises / fears [71]. These works motivate us to investigate similar differences in the SE domain. Our next research question investigates the use of various categories of sentiment words (Table II) based on the gender of a developer.

RQ3: *Do the categories of sentiment words used during code reviews vary based on the gender of a developer?*

D. Gender Vs. Emoticon

Emoticons are widely used in informal written communication and help the author expressing his/her feelings or mood. Recent studies suggest significant differences based on gender with females using emoticons more frequently than males [71], [73]. Therefore, we are interested to find out if these results are supported in a SE context.

RQ4: *Does the likelihood of using emoticons during code reviews depend on the gender of a developer?*

³<https://code.google.com/p/gerrit/>

⁴<http://phabricator.org/>

⁵<https://www.reviewboard.org/>

E. Gender vs. Expletive / Swear Word

Swearing, or the use of expletives, are perceived as intrinsically forceful or aggressive activities. Cultural stereotype as well as the association of expletives with ‘masculine identity’ [41] suggest a less likelihood of expletives from females than from males. However, recent studies on a contemporary culture [64], [57] did not find any significant differences. Our next question explores whether female developers follow a traditional stereotype or a contemporary culture [64].

RQ5: *Does the likelihood of using swear words /expletives during code reviews depend on the gender of a developer?*

IV. RESEARCH METHODOLOGY

The accurate identification of the gender of a contributor is not only essential but also challenging for this research. In the following subsections, we describe our data collection, gender resolution strategy, and sentiment analysis methodology.

A. Data Collection and Preparation

We used the Gerrit-Miner tool of [17] to mine completed code reviews of 12 popular OSS projects and stored the data in a MySQL database. Among the 12 projects, we excluded the six projects that did not satisfy either of the following two criteria: i) an open source project that mandates each and every change to be submitted for reviews on Gerrit; and ii) project contributors have performed at least 30,000 code reviews. The selected six projects had total 564,891 completed (i.e., ‘Merged’ or ‘Abandoned’) code reviews. Table I shows the list of projects.

A manual inspection of the comments posted by some accounts (e.g., ‘Qt Sanity Bot’ or ‘BuildBot’) suggested that those accounts were automated bots rather than humans. These accounts typically contain one of the following keywords: ‘bot’, ‘auto’, ‘CI’, ‘Jenkins’, ‘integration’, ‘build’, ‘hook’, ‘recheck’, ‘travis’, or ‘verifier’. Because we wanted only code reviews from actual reviewers, we excluded these bot accounts after a manual inspection had confirmed that the interactions were automatically generated. Following a similar approach as Bird *et al.* [14], we used the Levenshtein distance between two names to identify similar names. If our manual reviews of the associated accounts suggested that those belong to the same person we merged those to a single account.

In this study, we define a ‘*Non-casual developer*’ as a developer who has submitted at least five code changes for his/her project. Since our gender resolution strategy is time consuming, we only considered the non-casual developers in each project for our subsequent analyses. Column ‘Total devs.’ and column ‘non-casual devs.’ in Table I show the total number of developers and the number of non-casual developers in each of the six projects respectively. Although, the number of non-casual developers may be as low as 38% of total developers (i.e., Android), they contributed more than 95% of total code reviews in each of the six projects (column ‘% CR by non-casual developers’ in Table I). Our data preparation steps

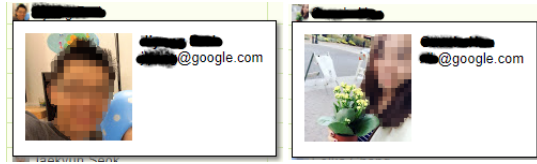


Fig. 1. Gender resolution using Gerrit profile pictures: 1) male (left), 2) female (right)

generated a list of 3,570 non-casual developers (Table I: Non-casual devs.) from the six projects.

B. Gender Resolution

We adopted a semi-automated gender resolution strategy using the genderComputer tool created by Vasilescu *et al.* [67] and modified by Terrell *et al.* [65] and followed the automated steps with manual validations using publicly available information on social networks. The genderComputer tool uses a database of 221,854 first names from 204 countries around the world and classifies each name belonging to one of the following four categories:

- 1) *Male-* names given to males.
- 2) *Female-* names given to females.
- 3) *Unisex* - names given to both males and females.
- 4) *None*- no entry in the database.

Based on the names of our 3,570 non-casual developers from the six projects, the genderComputer tool classified the contributors as following: 2,633 males, 325 females, 496 unisex, and 116 none. To ensure the accuracy of the identified genders, we adopted following five-step manual validation strategy for the 937 non-male (i.e., female, unisex, or none) contributors. We moved onto the next resolution step only if all the previous steps failed. If all the five manual validation steps were unsuccessful, which occurred for only 39 contributors ($\approx 1\%$), we excluded a developer from our subsequent analyses. Although, our gender resolution steps leverage only publicly available information, we got our research method reviewed and approved by our Institutional Review Board (IRB).

1) *Resolution using Gerrit Avatar:* Gerrit allows a user to include his/her picture in his/her profile. In the first step, we look into the Gerrit avatar of a user to determine his/her gender. Figure 1 shows examples of two Gerrit avatars. The avatar on the left indicates a male contributor and the avatar on the right indicates a female. However, some users’ Gerrit avatars were either empty or images that do not reveal their genders.

2) *Resolution using Google plus:* We searched the Google plus social network using an user’s email address. Based on Google plus search policy, if an user has associated his/her email address with a profile, a search based on that email address returns only that particular profile. Since the gender information of an user on Google plus is public for the majority of the users, a positive match based on an email search potentially could help finding the user’s gender (Figure 2).

TABLE I
PROJECT DEMOGRAPHICS

Project	Domain	Technology	Using Gerrit since	Requests mined*	Total devs.	Non-casual devs.	% CR by non-casual devs	% Female
Android	Mobile OS	C, C++, Java	October, 2008	81,137	2,589	981	95.6%	6.19%
Chromium OS	Desktop OS	C, C++	March, 2011	153,523	1,511	1,019	99.2%	8.74%
Couchbase	NoSQL database	C, C++	May, 2010	64,799	247	165	99.7%	9.69%
OmapZoom	Mobile Platform	C	February, 2009	35,973	604	439	95.9%	7.06%
OVirt	Virtualization	Java	October, 2011	73,523	345	220	99.4%	9.54%
Qt	UI framework	C, C++	May, 2011	155,936	1,598	746	94.9%	3.12%
*Mined during September, 2017				Total:	564,891	3,570		

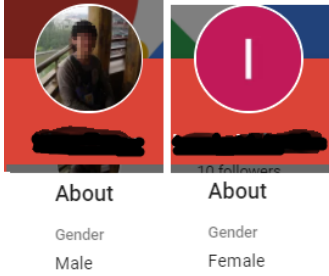


Fig. 2. Gender resolution using Google plus profile: 1) male (left), 2) female (right).

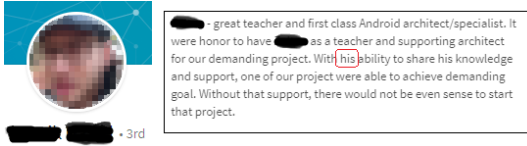


Fig. 3. Gender resolution: 1) using LinkedIn profile picture (left), 2) using received recommendations (right).

3) *Resolution using LinkedIn Profile*: For these users, we searched LinkedIn, a professional social network, with his/her full name and company information. For example, if a user’s name is ‘Kai Chen’ and his/her email address is ‘kai.chen@intel.com’, we searched using the term ‘Kai Chen + Intel’. If we found a positive match, we inspected the profile picture to determine his/her gender. However, if a user’s profile picture was invisible to us, we looked into the recommendations that he/she has received. Any gender specific pronouns (i.e., ‘he’, ‘she’, ‘his’, or ‘her’) in the recommendations revealed the user’s gender. Figure 3 shows examples of gender resolutions using the LinkedIn.

4) *Resolution using Facebook*: We used the same search term as used on the LinkedIn (‘full_name + company_name’) on Facebook. If a positive match was found, we inspected the profile pictures as well as gender specific pronoun in the phrase (‘To see what he/she shares..’) to determine a user’s gender.

5) *Resolution using Google Search*: If all the first four steps failed (< 3% overall), we searched on Google using ‘full_name + company_name’ to identify the profiles of a user on various other platforms (e.g., blog, presentation, video, Twitter, Github, and forums). If information obtained from

those platforms suggests a positive match, we inspected pictures or referring pronouns on those platforms to identify the gender of a user.

C. Sentiment Analysis

Since sentiment analysis tools built for other domains do not work well on a SE dataset [6], researchers have recently proposed several custom tools for the SE domain, such as SentiCR [6], Senti4SD [18], SentiStrength-SE [38], and SentiSE [36]. Among the four tools, we use SentiSE, since i) it has the largest training dataset of 13K SE interactions among the four tools; ii) SentiSE’s training dataset contains 2,800 code review comments, which are focused in this study; and iii) SentiSE boasts the highest accuracy (86.9%), the highest weighted kappa [21] (0.788), and the highest F-measure for all three classes i.e., positive (86.9%), neutral (89.0%), and negative (82.1%). SentiSE⁶ is open source and publicly available at:

<https://github.com/amiangshu/SentiSE>.

We use SentiWordNet 3.0 [11] to compile a list of words expressing sentiments. SentiWordNet is a collection of *synsets*, where each synset contains one or multiple words with the same meaning. SentiWordNet identifies each synset with a unique ID, a positive score, a negative score and a definition. SentiWordNet contains 14,021 words that can potentially express sentiments (i.e. have nonzero positive/negative scores). The objective of this research (i.e. RQ2) also requires grouping the sentiment words into categories. Among the existing categorizations, Arnold proposed the first scheme with eleven fundamental emotions [8]. Later, Plutchik [62] and Parrot [59] published reduced schemes with eight groups and six groups respectively. Combining the three aforementioned classifications, SentiSense [24] proposes 14 emotional categories and classifies 5,496 commonly used sentiment words from the SentiWordNet. Table II shows example words from the 14 SentiSense categories.

We leverage the list of emoticons and swear words / expletives compiled in SentiSE [36]. Our list of swear words / expletives include 84 commonly used ones from the english language and list of emoticons include total 107 emoticons with 46 indicating positive sentiments and the remaining 61 indicating negatives.

TABLE II
EXAMPLE WORDS FROM THE 14 SENTISENSE CATEGORIES

Category	Example words
Ambiguous	ironic, rare, strong, thrilling, uncommon
Anger	danger, disturbing, intolerable, troubling, worrisome
Anticipation	ambition, aspiration, fair, reasonable, sufficient
Calmness	innocent, patient, placid, resolute, smooth
Despair	despair, discourage, hopeless, pessimism, unsupportive
Disgust	confusion, disappointing, disgraceful, insane, mistake
Fear	disastrous, horrible, rage, scary, vicious
Hate	abysmally, crappy, rotten, shitty, worthless
Hope	credible, encouraging, exalted, optimistic, sublime
Joy	cheer, congratulate, fortuitous, jubilant, satisfactory
Like	advantage, awesome, excellent, great, magnificent
Love	adorable, kudos, lovely, marvelous, wonderful
Sadness	alas, cry, doomed, regret, unlucky
Surprise	amazed, astonishing, misleading, surprising, wonder

V. RESULTS

In this section, we present the results of our analyses to answer the five research questions introduced in Section III. Table III shows the results of Chi-Square tests for the five research questions to find statistical significance of the differences between male and female developers.

A. RQ1: Gender vs. Sentiment

Figure 4, shows the ratios of negative and positive review comments authored by male and female developers from the six projects. More than 85% code review comments in those projects are neutrals (i.e., do not express any sentiments). Among the review comments expressing sentiments, the ratios of negative sentiments are higher than positives for all six projects. For example, around 12% reviews authored by the male developers from Android expressed negative sentiments compared to less than 3% positives. Since the primary goal of code reviews is to identify mistakes in code, a higher ratio of negative sentiments than positives may not be surprising.

Five out of the six projects (i.e., except OmapZoom) indicate significantly (Table III: RQ1) higher likelihood of review comments authored by males expressing either negative or positive sentiments than those authored by females. For example, in Qt, around 4% of the review comments authored by males were positives compared to only 2% from females. Similarly, $\approx 12\%$ reviews from males were negatives compared to $\approx 6\%$ from females. We also found more than 90% review comments authored by female developers as neutrals.

Finding 1: Male developers were significantly more likely to author review comments expressing positive / negative sentiments than females.

⁶A publication detailing the design and evaluation of SentiSE is currently under review.

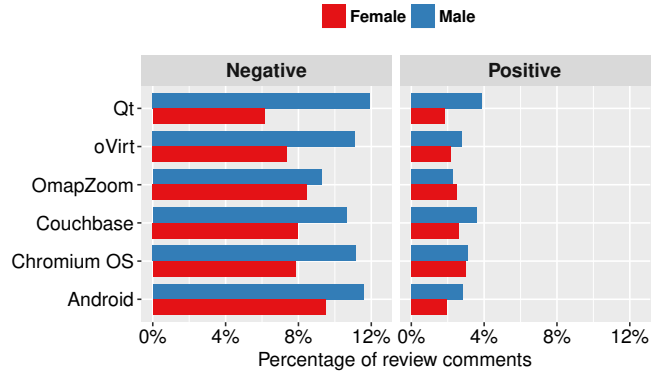


Fig. 4. Distribution of sentiments: Male vs. Female

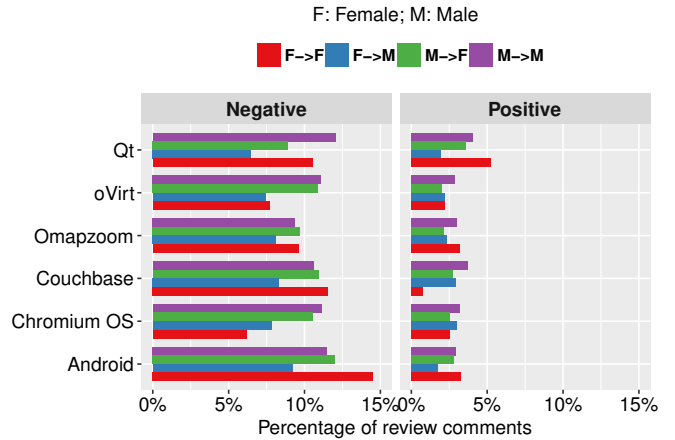


Fig. 5. Distribution of sentiments: Same gender vs. cross-gender

B. RQ2: Same Gender Vs. Cross-Gender Interaction

We computed the distributions (Figure 5) of negative and positive review comments during male \rightarrow male, male \rightarrow female, female \rightarrow male, and female \rightarrow female interactions. Our results suggest significant differences among those distributions for all the six projects (Table III: RQ2). Four out of the six projects (ie., except Couchbase and Chromium OS) show females more frequently expressing both positive and negative sentiments during their interactions with another female than during than interactions with a male. In Couchbase, females were more negative but less positive when communicating with another female. However, Chromium OS shows an exception with females expressing less sentiments to other females than to males.

For males, the results are mixed. In Qt, Chromium, and oVirt males expressed more sentiments, both positives and negatives, during their interactions with another male than during their interactions with a female. But male developers seem to be harsh to their female colleagues in Android, Couchbase, and Omapzoom, as males not only wrote negative reviews more frequently but also wrote positive reviews less frequently during their interactions with a female than during their interactions with another male.

TABLE III
RESULTS OF THE STATISTICAL TESTS FOR OUR FIVE RESEARCH QUESTIONS

Project	RQ1: Sentiment		RQ2: Cross-gender sentiment		RQ3: Types of sentiment words		RQ4: Emoticons		RQ5: Expletives / Swear words	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Android	27.98	< 0.01*	34.18	< 0.01*	249.62	< 0.01*	38.37	< 0.01*	5.70	0.02*
Chromium OS	182.06	< 0.01*	211.03	< 0.01*	361.35	< 0.01*	0.005	0.94	21.35	< 0.01*
Couchbase	20.74	< 0.01*	20.24	< 0.01*	347.02	< 0.01*	71.79	< 0.01*	4.07	0.04*
OmapZoom	2.27	0.32	14.15	0.02*	1467.4	< 0.01*	0.08	0.78	0.27	0.60
oVirt	188.81	< 0.01*	165.71	< 0.01*	225.87	< 0.01*	2810.2	< 0.01*	3.52	0.05*
Qt	98.98	< 0.01*	96.18	< 0.01*	418.52	< 0.01*	29.07	< 0.01*	4.12	0.04*

*-statistically significant

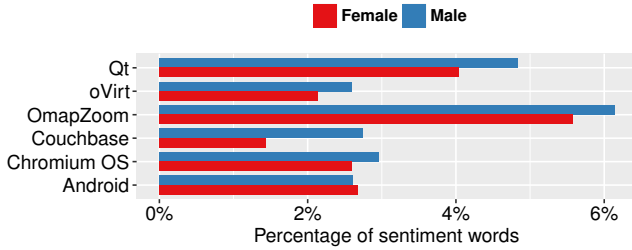


Fig. 6. Frequency of sentiment words: Male vs. Female

Finding 2: In five of the six projects, female developers were more likely to express sentiments to another female than to a male. However, in three out of the six projects, males were harsher to females by not only providing more negative reviews but also providing less positive encouragements.

C. RQ3: Gender Vs. Sentiment Word

We computed the frequencies of each word belonging to the *SentiSense* [24] among the reviews authored by both male and female developers. Among the 14 categories from *SentiSense*, words belonging to the ‘ambiguous’ category assume sentiment orientations based on its context. Moreover, words belonging to the ‘despair’ category were rare or even non-occurring among the review comments in our dataset. Therefore, we excluded these two categories from our analysis.

In five of the six projects (except Android) male developers used sentiment words more frequently than females (Figure 6). Figure 7 shows the occurrences of the 12 *SentiSense* categories per 100K words authored by male and female developers from the six projects. Since words from some categories were several times more frequent than the other categories, we plot those charts on a log scale. For example, males from Android used words from the ‘like’ category 1,510 times per 100k words, but the corresponding number for the words from the ‘fear’ category was only 23. We also noticed significant differences between males and females in using *SentiSense* words (Table III: RQ3).

In three of the six projects (i.e., except Android, Couchbase, and OmapZoom), males were more likely to use words

belonging to strong sentiment categories such as ‘Surprise’, ‘Sadness’, ‘Love’, ‘Joy’, ‘Hate’, ‘Disgust’, ‘Fear’, and ‘Anger’. The frequencies for the words belonging to mild sentiment categories such as ‘Like’, ‘Anticipation’, ‘Calmness’ and ‘Hope’ were very similar between males and females. Females from Android, Couchbase and Omapzoom were more frequently expressing ‘Anger’ and ‘Disgust’ possibly due to the negative attitudes from their male colleagues.

Finding 3: Male developers were significantly more likely to use sentiment words than females. Moreover, during sentiment expressions, females were less likely to use words expressing strong sentiments than males.

D. RQ4: Gender Vs. Emoticon

Figure 8 shows the distributions of positive / negative emoticons per 1000 review comments authored by male and female developers. In general, developers use positive emoticons far more frequently than negatives. The two ‘smiley face’ emoticons (i.e., ‘:)’ and ‘:-)’) accounted for more than 75% of the emoticon usage for each project. Among the negative emoticons, sad face ‘:(’ and tongue out ‘:p’ were the most frequent ones. We also noticed emoticons more frequent among comments from males than from females (except positive emoticons in Chromium OS). These differences between male and female developers in using emoticons are statistically significant for four out of the six projects (except Chromium OS and OmapZoom) as shown in Table III: RQ4. However, we did not notice any significant affinity towards a particular emoticon based on gender.

Finding 4: Contrary to prior results, female developers were less likely to use emoticons than males during code reviews.

E. RQ5: Gender Vs. Expletive / Swear Word

Figure 9 shows the distributions of expletives / swear words per 100k words authored by male and female developers. In general, the frequencies of expletives / swear words were very low, which may not be surprising since our dataset includes six of the top OSS projects. We also noticed male developers more frequently using expletives / swear words than females. These differences are also statistically significant for five out of the

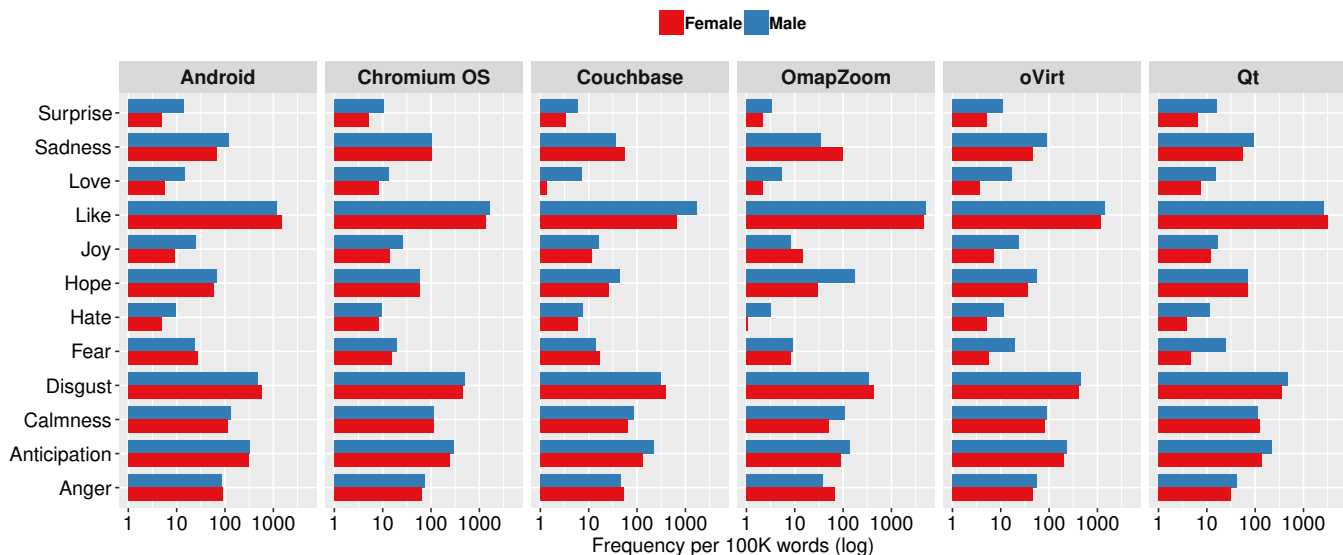


Fig. 7. Distribution of sentiment word categories from SentiSense: Male vs. Female

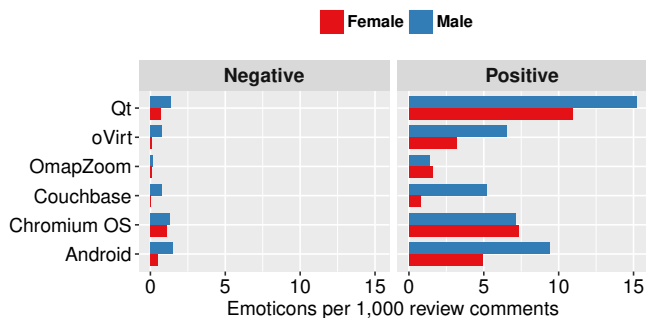


Fig. 8. Distribution of emoticons: Male vs. Female

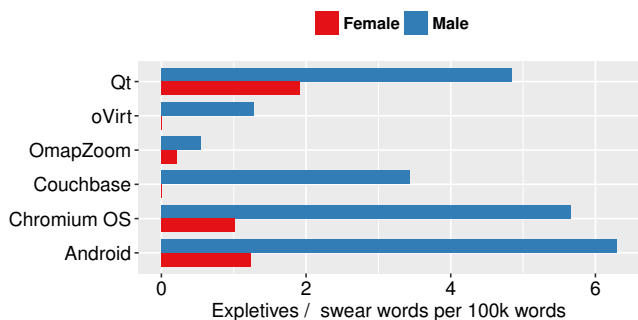


Fig. 9. Distribution of expletives / swear words: Male vs. Female

six projects (i.e., except OmapZoom) as shown in (Table III: RQ5). We also noticed ‘crap’, ‘damn’, and ‘screw’ as the most common expletives used by both males and females. However, some of the highly offensive expletives such as ‘bitch’, ‘bastard’, ‘fuck’, ‘jerk’ were used only by males.

Finding 5: *Female developers were significantly less likely to use expletives than males in five out of six projects. Even when female used expletives, they avoided certain highly offensive ones that males often used.*

VI. DISCUSSION AND IMPLICATIONS

In this section, we compare our results with prior results from other domains and discuss possible reasons and implications for our findings.

A. Verbal Abuses Towards Females

One of the primary goal of this study is to determine whether females are more frequent recipients of negative opinions than males. The results are mixed, since three out of the six projects indicate possible discrimination with males not only writing negative reviews more frequently to females but also withholding positive encouragements (Section V-B). Unsurprisingly, females from those projects were more frequently using words expressing ‘Anger’ or ‘Disgust’ than the males belonging to those projects (Section V-C) in their reviews. However, it is encouraging to see this trend as not a norm across all the six projects.

On the other hand, the results of RQ5 (Section V-E), which suggest more frequent expletives from the males than from the females, with some of expletives being highly offensive to females, are concerning. Some software projects, especially OSS projects have been guilty of fostering a ‘toxic culture’ for women, where inflammatory talk and aggressive posturing is acceptable within the norm of the community [53]. Some words often used in the mailinglists are insulting to women. Recently similar toxic culture has been also reported in Apple [54], Tesla [46] and UploadVR [40], where crude and sexist jokes often made female employees uncomfortable. A similar scenario seems to be present across all the six projects in our study.

B. Comparisons with Other Domains

While some of our results support prior findings in other areas, several of our results are contradicting. First, while women have been termed as more emotional than men in most of the prior studies [51], [66], [74], we find a completely opposite picture in the SE context. Second, Thelwal et al. [66], found male-female interactions to be slightly flirtatious, while female-female interactions to be transparently supportive. On the contrary, our findings indicate that females remain neutral while interacting with males, but males render more negative reviews to their female peers. Although females writing more positive comments to another female may hint support, females also wrote more negative reviews to another female. Third, prior studies found a tendency for women to report feeling stronger and longer emotions and to express them more clearly [28]. However, five out of our six projects indicate a contradicting picture with males using stronger emotions more frequently than females. Fourth, in terms of emoticon usage, Wolf found women more frequently using emoticons than men. While women used humors more frequently, emoticons to express teasing or sarcasm dominated in males' newsgroups [73]. However, our findings not only found less emoticons from females but also did not find any significant differences based on gender. Finally, although recent studies suggest [64], [57] females breaking traditional stereotype in the usage of expletives or swear words, we did not observe a similar trend in the SE context.

C. Non-expressiveness of Female Developers

One intriguing question gets raised by our results is, *why female developers do not express their sentiments in the SE context as they often do in social media*. While we do not have a definite answer, several factors, as discussed following may be contributing. Inclusive communities with more women can be more comfortable for women to ask questions or interact with each other. However women constitute less than 10% OSS developers [44]. A recent survey on Github developers also report profound gender imbalance with 95% males compared to only 3% females [4]. As a result, women may be misinterpreted in software engineering while expressing their feelings or sentiments. A woman who speaks up may be judged as aggressive in OSS projects [44] and therefore, in most of the cases, women keep themselves silent and do not speak up or show emotions. In another survey [33], female executives reported that their comments during heated discussions are misinterpreted and are perceived as emotional. Therefore, women were advised to back up their feelings with logic, specificity and facts.

The uneasiness or self-doubt and less confidence in female developers may also arise from the lack of female mentors or seniors in software engineering. For example, women held only 11% of top executive positions at Silicon Valley companies [42]. Moreover, women were three-and-half times more likely to holding junior positions, in spite of being equally capable as their male peers [7]. However, the presence of female role models, as recently observed in

Malaysia [50], may make the SE industry less associated with masculine characteristics and more appropriate for women. Sheryl Sandberg describes this issue as *"There aren't more women in tech because there aren't more women in tech.* [45]."

In a recent survey by Elephant in the Valley [69], 87% of the women working in the Silicon Valley reported demeaning comments from their male colleagues. More alarmingly, 60% of the women reported sexual harassments, which is 1.5 times higher than the percentage of women reporting the same in other areas (41%) [23]. Moreover, women are often questioned on their moral standing and face social consequences for their linguistic behavior. Using aggressive or harsh or forceful words is treated as infraction of their cultural and social stereotypes. Using expletives breaches behavioral compliance in some communities and society imposes more obligations on women in preserving social values and norms than men [64]. Due to these factors, female developers who do not feel comfortable at their work environments, may subdue their natural instincts [74] to express emotions.

D. Implications

In recent years the US government as well as IT organizations have taken numerous initiatives to increase the enrollments of females in computing education [1], [2], [3], which helped to improve the percentage female CIS graduates to 18% in 2016 [76] from 12% in 2006. However, retaining women in computing professions received very little attention [58]. A recent report suggests that 45% women who chose computing careers leave the field within ten years and that quit rate is more than twice as high for women than it is for men [9]. The primary reasons behind the attrition can be negative workplace experiences, lack of access to creative technical roles, and dissatisfaction with career prospects [9]. As a result, although the percentage of female graduates continued to increase during the last decade, the percentage of women holding computing jobs are declining [70]. According to a report from the US Department of Commerce, women accounted for 30% of the computing jobs in 2000 and that number fell to 27% in 2009 [13]. Another report by the "Girls Who Code" initiative estimated the percentage of women in computing jobs steady at 24% from 2011 to 2016 despite the growing ratio of female graduates and that number will likely reduce to 22% by year 2025 [2]. Being more frequent recipients of negative reviews than males (Section V-B) or encountering words that are demeaning to females (Section V-E) may be a factor behind the negative workplace experiences encountered by the female developers and may be contributing to the diminishing number of females in the SE domain.

A recent study [20] investigated the gender biasness in six of the largest natural science and engineering fields. The results found biological sciences, chemistry, mathematics, and statistics as more gender-balanced compared to computer science, engineering, and physics. According to Cheryan *et al.* the three dominant factors shaping preferences for one subfield over another are: i) insufficient early experience, ii) perceptions of a masculine culture, and iii) gender gaps

in self-efficacy. Some computing organizations foster ‘geek culture’ [48], which promotes certain stereotypes that do not fit well with most women. Therefore, women often feel as outsiders within those organizations. The characteristics and culture in communication among male and female counterparts in code reviews can represent the true nature of their inter relationship and the masculine culture in SE fields. So, this study will open up opportunity of research on gender bias in SE field from another point of view. Our study reveals the true nature of the behavioral differences between men and women in the SE field and further investigations can help overcoming diminishing number of females in this domain [2].

VII. THREATS

In the following subsections, we address three common types of threats to any empirical study.

A. Internal Validity

The primary threat to internal validity in this study is *project selection*. We included six publicly accessible OSS projects that practice tool-based code reviews supported by the same tool (i.e., Gerrit). Though, it is possible that projects supported by other code review tools (e.g., ReviewBoard, Github pull-based reviews, and Phabricator) could have behaved differently, we think this threat is minimal for four reasons: 1) all code review tools support the same basic purpose, i.e. detecting defects and improving the code, 2) the basic workflow (i.e. authors posting code, reviewers commenting about code snippets, and code requiring approval from reviewer before integration) of most of the code review tools are similar, 3) we did not use any Gerrit-specific feature/attributes in this study, and 4) sentiments expressed in review comments may not depend on any feature that is exclusive to Gerrit only. Therefore, we believe the project selection threat is minimal.

B. Construct Validity

The primary threat to construct validity is related to our gender resolution methodology based on the genderComputer tool, which has been used in several prior SE studies [68], [65]. Since we manually validated the tool’s classification for the 937 non-males, we believe that no male was misclassified as a female in our dataset. However, as we accepted genderComputer’s classifications for the 2,633 males, we are unable to make a similar claim that no female was misclassified as a male. To estimate, possible classification errors, we manually validated the genders of 200 randomly selected developers from the 2,633 males, using a similar methodology that we used for the non-males. Since our validation found only one female (0.5%) among these 200 developers, we do not think enough females were misclassified as males in our dataset to alter the results of our research questions.

We use SentiSE, a state-of-the-art sentiment analysis tool which is highly accurate ($\approx 87\%$). However, we cannot confidently claim that the misclassification ($\approx 13\%$) of SentiSE has not altered the results of our study. However, this would have been true only if there is a systematic relationship between the

comments that SentiSE incorrectly classifies and the gender of the author (if, for example, SentiSE incorrectly classifies comments written by females more frequently than comments written by males). We have no reason to believe that such a relationship exists, but have no empirical evidence.

C. External Validity

Although we analyzed a large number of code review requests from six popular and matured OSS projects, we cannot definitively establish that our sample is representative of the entire OSS population. Because OSS projects vary on characteristics like product, participant type, community structure, and governance, we cannot draw general conclusions about all OSS projects from this single study. To build reliable empirical knowledge, we need family of experiments [12] that include OSS projects of all types.

VIII. CONCLUSION

In this study, we explored the differences between men and women in using sentiment words, emoticons, and expletive during code reviews. This work is motivated by earlier findings where authors investigated existing gender bias in tech fields and differences of sentimental expression of men and women in other contexts such as social media. However, no prior research has explored the differences in opinions between men and women in the SE domain, and broadly in tech fields. Our results suggest that the likelihood of using sentiment words, emoticons, and expletives during code reviews vary based on the gender of a developer as women are less likely to use sentiment words / emoticons / expletives than men. We also investigated same-gender and cross-gender interactions and found female developers less frequently writing negative comments to their male counterparts. Yet, male developers from three out of the six projects were not only critical of their female counterparts but also withheld positive encouragements. Unsurprisingly, females from those projects were expressing ‘Anger’ and ‘Disgust’ more frequently than males. Our results also found males more frequently using expletives or words that are demeaning to females across all the six projects, supporting a ‘toxic culture’ as suggested by a prior study [53]. The results of this research can be used to analyze the ongoing gender issues in tech fields from different perspective. The diminishing trend of gender gap in other fields including automotive industry, biological sciences and mathematics [20] stems from the pursue of investigating the gender issues from the very beginning of education level to the professional environment. Therefore, our results will be interesting for the researchers to investigate the current socio-cultural norms against women in tech fields and for the professionals to create a workplace where every woman can feel confident, supported and safe to pursue their dreams. In a follow-up study, we plan to further investigate the outcomes of this study qualitatively and determine possible factors behind our results. We hope our results can motivate further research both to retain and to encourage more women in computing professions.

REFERENCES

- [1] (2017) Cs for all. [Online]. Available: <http://www.csforall.org/>
- [2] (2017) Girls who code. [Online]. Available: <https://girlswhocode.com/>
- [3] (2017) Nasa girls. [Online]. Available: <https://women.nasa.gov/about-nasa-girls/>
- [4] “Open source survey,” 2017, accessed = 2018-05-19. [Online]. Available: <http://opensourcesurvey.org/2017/>
- [5] J. Abbate, *Recoding gender: Women’s changing participation in computing*. MIT Press, 2012.
- [6] T. Ahmed, A. Bosu, A. Iqbal, and S. Rahimi, “Senticr: A customized sentiment analysis tool for code review interactions,” in *Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering*, 2017, pp. 106–111.
- [7] A. Al-Heeti, “Young women dominate in software, but still face setbacks,” 2018, accessed = 2018-05-19. [Online]. Available: <https://www.cnet.com/news/young-women-dominate-in-software-but-still-face-setbacks-says-hackerrank/>
- [8] M. Arnold, “Emotion and personality. volume i: Psychological aspects,” in *New York: Columbia University Press*, 1960.
- [9] C. Ashcraft, B. McLain, and E. Eger, “Women in tech: The facts,” *National Center for Women et Information Technology*, 2016.
- [10] A. Bacchelli and C. Bird, “Expectations, outcomes, and challenges of modern code review,” in *Proceedings of the 2013 International Conference on Software Engineering*. IEEE Press, 2013, pp. 712–721.
- [11] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining,” in *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, 2010, pp. 2200–2204.
- [12] V. R. Basili, F. Shull, and F. Lanubile, “Building knowledge through families of experiments,” *IEEE Transactions on Software Engineering*, vol. 25, no. 4, pp. 456–473, 1999.
- [13] D. Beede, T. Julian, D. Langdon, G. McKittrick, B. Khan, and M. Doms, “Women in stem: A gender gap to innovation. esa issue brief# 04-11.” *US Department of Commerce*, 2011.
- [14] C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan, “Mining email social networks,” in *Proceedings of the 2006 international workshop on Mining software repositories*. ACM, 2006, pp. 137–143.
- [15] S. Bird, “NLTK: The Natural Language Toolkit,” in *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics, 2006, pp. 69–72.
- [16] A. Bosu, J. C. Carver, C. Bird, J. Orbeck, and C. Chockley, “Process Aspects and Social Dynamics of Contemporary Code Review: Insights from Open Source Development and Industrial Practice at Microsoft,” *IEEE Transactions on Software Engineering*, vol. 1, no. 99, pp. 1–1, 2017.
- [17] A. Bosu and J. C. Carver, “Impact of peer code review on peer impression formation: A survey,” in *2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2013, pp. 133–142.
- [18] F. Calefato, F. Lanubile, F. Maiorano, and N. Novielli, “Sentiment polarity detection for software development,” *Empirical Software Engineering*, pp. 1–31, 2017.
- [19] K. Campbell and A. Mínguez-Vera, “Gender diversity in the boardroom and firm financial performance,” *Journal of business ethics*, vol. 83, no. 3, pp. 435–451, 2008.
- [20] S. Cheryan, S. A. Ziegler, A. K. Montoya, and L. Jiang, “Why are some stem fields more gender balanced than others?” *Psychological Bulletin*, vol. 143, no. 1, pp. 1–35, 2017.
- [21] J. Cohen, “Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.” *Psychological bulletin*, vol. 70, no. 4, p. 213, 1968.
- [22] J. Damore. Google’s ideological echo chamber: How bias clouds our thinking about diversity and inclusion. [Online]. Available: <https://gizmodo.com/exclusive-heres-the-full-10-page-anti-diversity-screed-1797564320>
- [23] A. Das, “Sexual harassment at work in the united states,” *Archives of sexual behavior*, vol. 38, no. 6, pp. 909–921, 2009.
- [24] J. C. de Albornoz, L. Plaza, and P. Gervas, “Sentsense: An easily scalable concept-based affective lexicon for sentiment analysis,” in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, 2012.
- [25] L. DiDio, “Look out for techno-hazing,” vol. 31, no. 39, pp. 72–73, 2017.
- [26] N. L. Ensmenger, *The computer boys take over: Computers, programmers, and the politics of technical expertise*. MIT Press, 2012.
- [27] R. Feldman, “Techniques and applications for sentiment analysis,” *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [28] A. H. Fischer and A. S. Manstead, “The relation between gender and emotions in different cultures,” *Gender and emotion: Social psychological perspectives*, vol. 1, pp. 71–94, 2000.
- [29] D. Garcia, M. S. Zanetti, and F. Schweitzer, “The role of emotions in contributors activity: A case study on the Gentoo community,” in *Cloud and Green Computing (CGC), 2013 Third International Conference on*. IEEE, 2013, pp. 410–417.
- [30] R. A. Ghosh, R. Glott, B. Krieger, and G. Robles, “Free/libre and open source software: Survey and study,” 2002.
- [31] E. Guzman, D. Azócar, and Y. Li, “Sentiment analysis of commit comments in github: an empirical study,” in *Proceedings of the 11th Working Conference on Mining Software Repositories*. ACM, 2014, pp. 352–355.
- [32] E. Guzman and B. Bruegge, “Towards emotional awareness in software development teams,” in *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*, ser. ESEC/FSE 2013, 2013, pp. 671–674.
- [33] K. Heath and J. Flynn, “How women can show passion at work without seeming ‘emotional’,” 2015, accessed = 2018-05-19. [Online]. Available: <https://hbr.org/2015/09/how-women-can-show-passion-at-work-without-seeming-emotional>
- [34] C. Herring, “Does diversity pay?: Race, gender, and the business case for diversity,” *American Sociological Review*, vol. 74, no. 2, pp. 208–224, 2009.
- [35] S. A. Hewlett, L. Sherbin, F. Dieudonne, C. Fargnoli, and C. Fredman, “Athena factor 2.0: Accelerating female talent in science, engineering & technology,” *Center for Talent Innovation 2014*, 2014.
- [36] T. Hossain, A. Kohli, and A. Bosu, “Sentsense: Sentiment analysis for software engineering interactions,” in *Under Submission to the Automated Software Engineering Journal*, 2018, p. TBD. [Online]. Available: <https://github.com/amiangshu/SentiSE>
- [37] M. R. Islam and M. F. Zibrán, “Exploration and exploitation of developers’ sentimental variations in software engineering,” *International Journal of Software Innovation (IJSI)*, vol. 4, no. 4, pp. 35–55, 2016.
- [38] —, “Leveraging automated sentiment analysis in software engineering,” in *Proceedings of the 14th International Conference on Mining Software Repositories*. IEEE Press, 2017, pp. 203–214.
- [39] T. James, M. Galster, K. Blincoe, and G. Miller, “What is the perception of female and male software professionals on performance, team dynamics and job satisfaction?: insights from the trenches,” in *Proceedings of the 39th International Conference on Software Engineering: Software Engineering in Practice Track*, 2017, pp. 13–22.
- [40] M. Kendall. (2017) Lawsuit describes vr startup’s office ‘kink room,’ parties ‘rife with sexual impropriety’. [Online]. Available: <https://tinyurl.com/y8fbuuo5>
- [41] V. d. Klerk, “Expletives: men only?” *Communications Monographs*, vol. 58, no. 2, pp. 156–169, 1991.
- [42] M. Kosoff, “Here’s evidence that it’s still not a great time to be a woman in silicon valley,” 2015, accessed = 2018-05-25.
- [43] R. Lakoff, *Language and woman’s place*, ser. Harper colophon books. Harper & Row, 1975. [Online]. Available: <https://books.google.com/books?id=0dFoAAAAIAAJ>
- [44] R. Lakshan, “The trouble with being a woman in the world of free and open source software,” 2015, accessed = 2018-05-19.
- [45] J. Lansing, “Women in tech: don’t even try to fit in a man’s world,” 2015, accessed = 2018-05-19. [Online]. Available: <http://fortune.com/2015/08/08/jane-lansing-women-in-tech/>
- [46] S. Levin. (2017) Female engineer sues tesla, describing a culture of ‘pervasive harassment’. [Online]. Available: <https://www.theguardian.com/technology/2017/feb/28/tesla-female-engineer-lawsuit-harassment-discrimination>
- [47] B. Liu and L. Zhang, “A survey of opinion mining and sentiment analysis,” in *Mining text data*. Springer, 2012, pp. 415–463.
- [48] J. Margolis and A. Fisher, *Unlocking the clubhouse: Women in computing*. MIT press, 2003.
- [49] A. Marwick, “Silicon valley and the social media industry,” *Sage Handbook of Social Media*. London: Sage, 2017.
- [50] U. Mellström, “The intersection of gender, race and cultural boundaries, or why is computer science in malaysia dominated by women?” *Social Studies of Science*, vol. 39, no. 6, pp. 885–907, 2009.

- [51] C. S. Montero, M. Munezero, and T. Kakkonen, "Investigating the role of emotion-based features in author gender classification of text," in *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 2014, pp. 98–114.
- [52] A. Murgia, P. Tourani, B. Adams, and M. Ortu, "Do developers feel emotions? an exploratory analysis of emotions in software artifacts," in *Proceedings of the 11th Working Conference on Mining Software Repositories*. ACM, 2014, pp. 262–271.
- [53] D. Nafus, J. Leach, and B. Krieger, "Gender: Integrated report of findings," *FLOSSPOLs, Deliverable D*, vol. 16, 2006.
- [54] u. . h.-s.-w.-r. Nathan Ingraham. (2017) Report: Apple is a sexist, toxic work environment.
- [55] M. Ortu, B. Adams, G. Destefanis, P. Tourani, M. Marchesi, and R. Tonelli, "Are bullies more productive?: empirical study of affectiveness vs. issue fixing time," in *Proceedings of the 12th Working Conference on Mining Software Repositories*. IEEE Press, 2015, pp. 303–313.
- [56] M. Ortu, G. Destefanis, S. Counsell, S. Swift, M. Marchesi, and R. Tonelli, "How diverse is your team? investigating gender and nationality diversity in github teams," *PeerJ Preprints*, vol. 4.
- [57] M. Ott, "Tweet like a girl: A corpus analysis of gendered language in social media," Yale University, Tech. Rep., 2016.
- [58] A. Panteli, J. Stack, M. Atkinson, and H. Ramsay, "The status of women in the uk it industry: an empirical study," *European Journal of Information Systems*, vol. 8, no. 3, pp. 170–182, 1999.
- [59] W. G. Parrott, *Emotions in social psychology: Essential readings*. Psychology Press, 2001.
- [60] K. Platman and P. Taylor, "Workforce ageing in the new economy: a comparative study of information technology employment," *University of Cambridge*, 2004.
- [61] D. Pletea, B. Vasilescu, and A. Serebrenik, "Security and emotion: sentiment analysis of security discussions on github," in *Proceedings of the 11th Working Conference on Mining Software Repositories*. ACM, 2014, pp. 348–351.
- [62] R. Plutchik, "Emotion: Theory, research, and experience: Vol. 1. theories of emotion," in *A general Psychoevolutionary Theory of Emotion*, 1980, pp. 3–33.
- [63] G. Robles, L. A. Reina, J. M. González-Barahona, and S. D. Domínguez, *Women in Free/Libre/Open Source Software: The Situation in the 2010s*, 2016.
- [64] K. Stapleton, "Gender and swearing: A community practice," *Women and Language*, vol. 26, no. 2, p. 22, 2003.
- [65] J. Terrell, A. Kofink, J. Middleton, C. Rainear, E. Murphy-Hill, C. Parnin, and J. Stallings, "Gender differences and bias in open source: Pull request acceptance of women versus men," *PeerJ Computer Science*, vol. 3, p. e111, 2017.
- [66] M. Thelwall, D. Wilkinson, and S. Uppal, "Data mining emotion in social network communication: Gender differences in myspace," *Journal of the Association for Information Science and Technology*, vol. 61, no. 1, pp. 190–199, 2010.
- [67] B. Vasilescu, A. Capiluppi, and A. Serebrenik, "Gender, representation and online participation: A quantitative study," *Interacting with Computers*, vol. 26, no. 5, pp. 488–511, 2014.
- [68] B. Vasilescu, D. Posnett, B. Ray, M. G. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov, "Gender and tenure diversity in github teams," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 3789–3798.
- [69] T. Vassallo, E. Levy, M. Madansky, H. Mickell, B. Porter, and M. Leas. (2017) Elephant in the valley. [Online]. Available: <https://www.elephantinthevalley.com/>
- [70] A. Vitores and A. Gil-Jurez, "The trouble with 'women in computing': a critical examination of the deployment of research on the gender gap in computer science," *Journal of Gender Studies*, vol. 25, no. 6, pp. 666–680, 2016.
- [71] S. Volkova, "Predicting user demographics, emotions and opinions in social networks," 2016.
- [72] A. Wilhelm. (2014) Julie ann horvath describes sexism and intimidation behind her github exit. [Online]. Available: <http://tcrn.ch/1m8Sf7w>
- [73] A. Wolf, "Emotional expression online: Gender differences in emoticon use," *CyberPsychology & Behavior*, vol. 3, no. 5, pp. 827–833, 2000.
- [74] V. L. Zammuner, "Mens and womens lay theories of emotion," *Gender and emotion: Social psychological perspectives*, pp. 48–70, 2000.
- [75] S. Zweben, "2006-2007 taulbee survey," *Computing Research News*, vol. 20, no. 5, 2008.
- [76] S. Zweben and B. Bizot, "2016 taulbee survey," *Computing Research News*, vol. 29, no. 5, 2017.