

Investigating the Effects of Gender Bias on GitHub

► Mounika Kari



Authors: Nasif Imtiaz , Justin Middleton , Joymallya Chakraborty, Neill Robson, Gina Bai, Emerson Murphy-Hill

Abstract



- Many software development firms emphasize diversity, including gender diversity, although the sector is still dominated by males.
- Gender bias is one explanation for this lack of diversity.
- The implications of that prejudice were investigated in this research by utilizing an existing framework drawn from the gender studies field.
- Four primary impacts of the framework were considered by providing assumptions about how they could emerge on GitHub.
- The findings reveal that the consequences of gender prejudice on the GitHub platform are mainly unnoticeable, yet there are still signs of women focusing their work on fewer locations and being more restricted in communication than males.

Introduction

- Gender diversity is a strong predictor of software development productivity.
- Unfortunately, promoting gender diversity in software engineering teams is difficult due to the low number of female computer scientists.
- Gender prejudice is one of the causes behind women's under representation in software engineering.
- To address the issue of gender bias in software development, we must first comprehend it and its consequences.
- For that, under the Williams and Dempsey's framework, four effects were examined
 - Prove-It-Again
 - Tightrope
 - Maternal wall
 - Tug of War

Related Work

- **Davison and colleagues' meta-analysis** - sex-type of a job affects employees' performance ratings and favors the sex it is typecast to
- **Roth and colleagues' meta-analysis** – women generally score higher in job performance ratings, although they still lag behind men in promotions
- **Heilman** - explains how gender stereotypes lead to biased judgments against women and impede their career progress
- **Vasilescu and colleagues** – women disengage from the platform sooner than men, even though they have similar activity levels
- **Nafus**- despite its reputation for openness and democracy, the complicated construct of "openness" in open-source software development results in a male monopoly.

General Methodology.

- **Research Setting: GitHub**

- a collaborative software development platform
- Each user gets a profile page where they may optionally add personal information and their project contribution history is automatically updated.

- **Dataset**

- builds on GHTorrent data mined from public GitHub data from June 7, 2010 to April 1, 2015
- The final dataset that used contained 152,534 PRs from 20,926 women and 3,135,384 PRs from 308,062 men

- **Measures and Measure Validation**

- New quantitative proxy measures adapted for the context of GitHub
- Construct validity of these measures were increased by
 - providing manual validation by inspection
 - triangulating on each bias effect by using multiple hypotheses



PROVE-IT-AGAIN

- When a member of a group who does not conform to preconceptions is held to a higher standard than those who do conform to stereotypes and must thus give more proof to establish competence.
- Work by contributors whose perceived gender is male is favored when the quality of work is comparable.
- This effect is especially pronounced in occupations that are traditionally assigned to males, such as software development.



Hypotheses considered & methodology used

1. Women provide longer pull request descriptions than men.
 - Length was measured as the character count of a PR description
2. Women's pull requests generate more discussion, receive more change suggestions, and take more time to get accepted.
 - Analysis was restricted to outsider pull requests from users whose gender is likely visible to other users
 - Acceptance time calculated as the time difference between when the pull request was first opened and its last merge state
3. Women put more signals of competence on their profile than men.
 - The proportion of men and women who have put information in their bio, company, and website URL field on their profile was analyzed
4. Women's pull requests remain more concentrated in fewer projects and fewer organizations than men's.
 - Each user's pull request concentration, across projects and organizations was measured

CHARACTERISTICS OF PRS FOR WOMEN AND MEN

| | Women | Men |
|---|-------|-------|
| No Description Provided (merged) | 47.4% | 29.9% |
| No Description Provided (non-merged) | 42.8% | 25.0% |
| Median character count of provided description (merged) | 103 | 164 |
| Median character count of provided description (non-merged) | 118 | 193 |
| PRs with review comments (merged) | 7.0% | 7.0% |
| PRs with review comments (non-merged) | 11.0% | 8.2% |
| PRs with review commits (merged) | 9.0% | 9.7% |
| PRs with review commits (non-merged) | 11.1% | 11.2% |
| Median hours for PR merge | 1.4 | 6.0 |

- ▶ Hypothesis 1 is not supported
- ▶ Hypothesis 2 is not supported

GITHUB PROFILE CHARACTERISTICS FOR WOMEN AND MEN

| | Women | Men |
|--|--------|---------|
| Total Profiles | 35 676 | 529 253 |
| Bio Provided | 7.3% | 11.7% |
| Median Character Count of Provided Bio | 37 | 37 |
| URL Provided | 17.9% | 31.2% |
| Company Provided | 18.7% | 29.7% |

- ▶ Hypothesis 3 is not supported



PULL REQUEST CONCENTRATION FOR WOMEN AND MEN

| | | | Women | Men |
|--|-----------------------|--------|-------|------|
| Concentration of PRs | Across Projects* | Mean | 2.9 | 2.4 |
| | | Median | 1.0 | 1.0 |
| | Across Organizations* | Mean | 5.8 | 4.8 |
| | | Median | 2.0 | 1.5 |
| Concentration of Net Lines of Code Changed | Across Projects* | Mean | 3418 | 2175 |
| | | Median | 26 | 25 |
| | Across Organizations* | Mean | 5258 | 2681 |
| | | Median | 47 | 29 |

- ▶ Hypothesis 4 is supported

Can be concluded that while women do not provide more evidence to demonstrate competence, they do concentrate their work across fewer projects and organizations, compared to men

TIGHTROPE

- Refers to the narrow band of socially acceptable behavior for women.
- Hypotheses considered
 1. Women avoid showing explicit politeness more than men.
 2. Women avoid profane words more than men.
 3. Women are more neutral in showing sentiment than men on GitHub.
 4. Compared to men, women use more sentiment-neutral ideograms.
 5. Women are more likely than men to avoid showing stereotypical masculine and stereotypical feminine traits.



TIGHTROPE ANALYSES FOR WOMEN AND MEN

| | Women | Men |
|--|-------|-------|
| Median Neutrality Rate for Politeness* | 66.0% | 62.2% |
| Percentage of Users never using Profanity* | 77.6% | 64.5% |
| Median Neutrality Rate for Sentiment* | 68.6% | 64.3% |
| Median Neutrality Rate in use of Ideogram with Sentiment* | 95.2% | 94.4% |
| Neutrality Rate in showing stereotypical Gendered Behavior | 71.5% | 73.0% |

- ▶ Hypothesis 1,2,3, and 4 are supported
- ▶ Hypothesis 5 is not supported.

The data suggests the Tightrope effect is largely evident, for which women tend to be more restrained than men in general.



MATERNAL WALL

- The bias against mothers wherein their coworkers perceive a choice without compromise
- Hypotheses considered
 1. The proportion of women who display that they are parents on GitHub is lower than the corresponding proportion of men.
- Evaluated by measuring whether women are less likely to include children in their profile pictures.



| | | Pictures with Children | | |
|-------|-------|------------------------|---------|------|
| | | Only G+ | Only GH | Both |
| Men | 1,547 | 35 | 6 | 5 |
| Women | 1,582 | 23 | 5 | 1 |

Hypothesis is not supported; we did not find evidence that women avoid posting pictures of children at a statistically significantly higher rates than men.

TUG OF WAR

- suggests that in heavily competitive environments, women sometimes discourage other women because either they doubt the competency of other women, or they think encouraging others may increase the level of competition
- related to the so-called “queen bee” syndrome
- Hypotheses considered
 1. Women are less likely to accept the pull requests created by other women, compared to pull requests created by men.
 2. Compared to pull requests made by men, pull requests made by women generate more discussions, receives more change suggestions, and take more time to be evaluated when reviewed by a woman
- PRs coming from identifiable men and women were analyzed, for which the PRs were reviewed by a woman.

PULL REQUESTS REVIEWED BY MEN AND WOMEN

| Reviewer | Creator | PR Acceptance Rate |
|----------|---------|--------------------|
| Women | Women | 88.4% |
| Women | Men | 88.0% |
| Men | Women | 84.0% |
| Men | Men | 80.1% |

PUSH-BACK AGAINST PRS BY WOMEN REVIEWERS

| | PR creator's Gender | |
|---------------------------------------|---------------------|------|
| | Women | Men |
| PRs with review comments (merged) | 4.0% | 5.8% |
| PRs with review comments (non-merged) | 8.1% | 7.9% |
| PRs with review commits (merged) | 6.2% | 7.8% |
| PRs with review commits (non-merged) | 15.7% | 8.7% |
| Median hours for PR merge | 0.2 | 0.9 |



- ▶ Hypothesis 1 had conflicting support and Hypothesis 2 is not supported. Therefore, substantial evidence for the Tug of War effect, where women would discourage other women were not found.

- No visible signs of women providing more evidence to demonstrate their competence than men.
- The result we observed can be an indirect outcome of the Prove-It-Again effect which leads women to prefer work in familiar places to avoid the Prove-It-Again cost in new projects and teams.
- Another explanation for the Prove-It-Again result could be that women, although lower in number in GitHub, are more competent than men overall.



- Hypotheses on Tightrope effects were largely supported. Interventions like,
 - Codes of conduct
 - Behavioral bots
 - Possess on tightrope effects in open-source help mitigate the tightrope effect.
- Maternal Wall was not supported.
 - One possible explanation is that mothers may not expect prejudice in virtual employment because it can be done anywhere, unlike labor in traditional offices. As a result, profile photographs may not be regarded as a sufficiently clear indicator of a user's maternal status.
- Hypotheses about the Tug of War effect were mostly not confirmed.
 - Although Hypothesis 1 is validated, study is limited to only insiders, there may be additional confounding variables at work when the data set shrinks due to the numerous filters applied throughout this research.

THREATS TO VALIDITY

- Whether the gender data is accurate.
- Some of the underlying data might have changed while carrying out the collection
- Inaccuracies in the tools we used might yield inaccurate results.
- Results are also threatened by construct validity.
- External validity

CONCLUSION



- Cues on GitHub that are comparable to those in physical workplaces were identified and investigated and may reflect the presence of gender bias effects.
- In some cases, evidence was found of hypothesized effects, whereas in other cases hypotheses were not supported.
- Women's underrepresentation in software development is concerning. It is essential to look for the possible effects of such bias and how we might address it

Your Thoughts

- Components on GitHub may be indicative of sex discrimination effects.
- The underrepresentation of women in software engineering is problematic and it is critical to investigate the potential consequences of such prejudice and how we could remedy it.



THANK YOU!