Gender Bias in Open Source: Pull Request Acceptance of Women Versus Men

Josh Terrell[1], Andrew Kofink[2], Justin Middleton[2], Clarissa Rainear[2], Emerson Murphy-Hill[2*], Chris Parnin[2]

[1]Department of Computer Science, Cal Poly, San Luis Obispo, USA
[2]Department of Computer Science, North Carolina State University, USA

*To whom correspondence should be addressed; E-mail: emerson@csc.ncsu.edu

Biases against women in the workplace have been documented in a variety of studies. This paper presents the largest study to date on gender bias, where we compare acceptance rates of contributions from men versus women in an open source software community. Surprisingly, our results show that women's contributions tend to be accepted *more* often than men's. However, when a woman's gender is identifiable, they are rejected more often. Our results suggest that although women on GitHub may be more competent overall, bias against them exists nonetheless.

Introduction

In 2012, a software developer named Rachel Nabors wrote about her experiences trying to fix bugs in open source software.[1] Nabors was surprised that all of her contributions were rejected by the project owners. A reader suggested that she was being discriminated against because of

[1]http://rachelnabors.com/2012/04/of-github-and-pull-requests-and-comics/

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

her gender.

Research suggests that, indeed, gender bias pervades open source. The most obvious illustration is the underrepresentation of women in open source; in a 2013 survey of the more than 2000 open source developers who indicated a gender, only 11.2% were women (*1*). In Vasilescu and colleagues' study of Stack Overflow, a question and answer community for programmers, they found "a relatively 'unhealthy' community where women disengage sooner, although their activity levels are comparable to men's" (*2*). These studies are especially troubling in light of recent research which suggests that diverse software development teams are more productive than homogeneous teams (*3*).

This article presents an investigation of gender bias in open source by studying how software developers respond to *pull requests*, proposed changes to a software project's code, documentation, or other resources. A successfully accepted, or 'merged,' example is shown in Figure 1. We investigate whether pull requests are accepted at different rates for self-identified women compared to self-identified men. For brevity, we will call these developers 'women' and 'men,' respectively. Our methodology is to analyze historical GitHub data to evaluate whether pull requests from women are accepted less often. While other open source communities exist, we chose to study GitHub because it is the largest (*4*), claiming to have over 12 million collaborators across 31 million software repositories.[2]

The main contribution of this paper is an examination of gender bias in the open source software community, enabled by a novel gender linking technique that associates more than 1.4 million community members to self-reported genders. To our knowledge, this is the largest scale study of gender bias to date.

[2]https://github.com/about/press

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

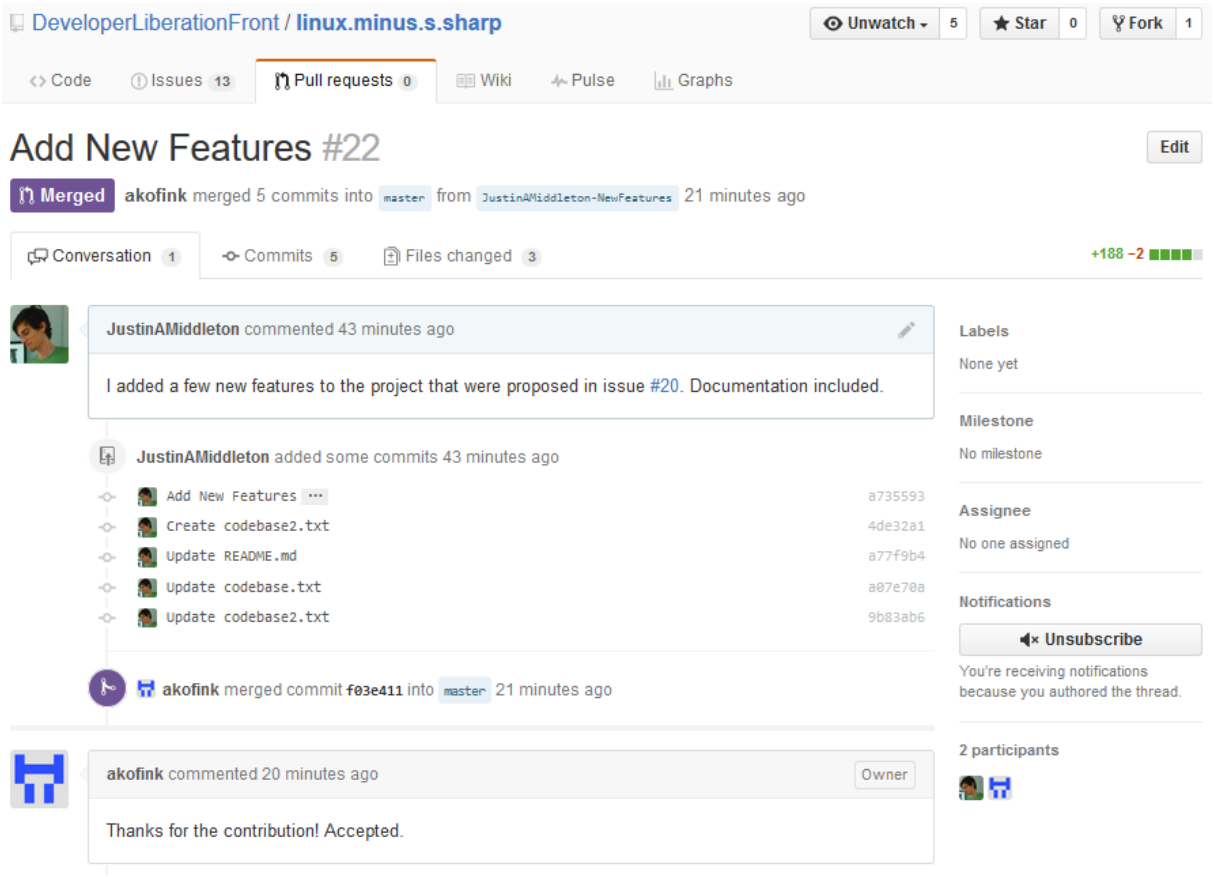

Figure 1: GitHub user 'JustinAMiddleton' makes a pull request; the repository owner 'akofink' accepts it by merging it. The changes proposed by JustinAMiddleton are now incorporated into the project.

Related Work

A substantial part of activity on GitHub is done in a professional context, so studies of gender bias in the workplace are relevant. Because we cannot summarize all such studies here, we instead turn to Davison and Burke's meta-analysis of 53 papers, each studying between 43 and 523 participants, finding that male and female job applicants generally received lower ratings for opposite-sex-type jobs (e.g., nurse is a female sex-typed job, whereas carpenter is male sex-typed) (*5*).

The research described in Davison and Burke's meta-analysis can be divided into experi-

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

ments and field studies. Experiments attempt to isolate the effect of gender bias by controlling for extrinsic factors, such as level of education. For example, Knobloch-Westerwick and colleagues asked 243 scholars to read and evaluate research paper abstracts, then systematically varied the gender of each author; overall, scholars rated papers with male authors as having higher scientific quality (*6*). In contrast to experiments, field studies examine existing data to infer where gender bias may have occurred retrospectively. For example, Roth and colleagues' meta-analysis of such studies, encompassing 45,733 participants, found that while women tend to receive better job performance ratings than men, women also tend to be passed up for promotion (*7*).

Experiments and retrospective field studies each have advantages. The advantage of experiments is that they can more confidently infer cause and effect by isolating gender as the predictor variable. The advantage of retrospective field studies is that they tend to have higher ecological validity because they are conducted in real-world situations. In this paper, we use a retrospective field study as a first step to quantify the effect of gender bias in open source.

Several other studies have investigated gender in the context of software development. Burnett and colleagues analyzed gender differences in 5 studies that surveyed or interviewed a total of 2991 programmers; they found substantial differences in software feature usage, tinkering with and exploring features, and in self-efficacy (*8*). Arun and Arun surveyed 110 Indian software developers about their attitudes to understand gender roles and relations but did not investigate bias (*9*). Drawing on survey data, Graham and Smith demonstrated that women in computer and math occupations generally earn only about 88% of what men earn (*10*). Lagesen contrasts the cases of Western versus Malaysian enrollment in computer science classes, finding that differing rates of participation across genders results from opposing perspectives of whether computing is a "masculine" profession (*11*). The present paper builds on this prior work by looking at a larger population of developers in the context of open source communities.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

Some research has focused on differences in gender contribution in other kinds of virtual collaborative environments, particularly Wikipedia. Antin and colleagues followed the activity of 437 contributors with self-identified genders on Wikipedia and found that, of the most active users, men made more frequent contributions while women made larger contributions (*12*).

There are two gender studies about open source software development specifically. The first study is Nafus' anthropoligical mixed-methods study of open source contributors, which found that "men monopolize code authorship and simultaneously de-legitimize the kinds of social ties necessary to build mechanisms for women's inclusion" (*13*). The other is Vasilescu and colleagues' study of 4,500 GitHub contributors, where they inferred the contributors' gender based on their names and locations (and validated 816 of those genders through a survey); they found that gender diversity is a significant and positive predictor of productivity (*3*). Our work builds on this by investigating bias systematically and at a larger scale.

General Methodology

Our main research question was

> To what extent does gender bias exist among people who judge GitHub pull requests?

To answer this question, we approached the problem by examining whether men and women are equally likely to have their pull requests accepted on GitHub, then investigated why differences might exist. While the data analysis techniques we used were specific to each approach, there were several commonalities in the data sets that we used, as we briefly explain below and in more detail in the Material and Methods appendix.

We started with the GHTorrent (*14*) dataset from April 1st, 2015, which contains public data pulled from GitHub about users, pull requests, and projects. We then augmented this GHTorrent

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

data by mining GitHub's webpages for information about each pull request status, description, and comments.

GitHub does not request information about users' genders. While previous approaches have used gender inference (*2, 3*), we took a different approach – linking GitHub accounts with social media profiles where the user has self-reported gender. Specifically, we extract users' email addresses from GHTorrent, look up that email address on the Google+ social network, then, if that user has a profile, extract gender information from these users' profiles. Out of 4,037,953 GitHub user profiles with email addresses, we were able to identify 1,426,121 (35.3%) of them as men or women through their public Google+ profiles. We are the first to use this technique, to our knowledge.

As an aside, we believe that our gender linking approach raises privacy concerns, which we have taken several steps to address. First, this research has undergone human subjects IRB review,[3] research that is based entirely on publicly available data. Second, we have informed Google about our approach to determine whether they believe that it's a privacy violation of their users to be able to link email address to gender; they responded that it's consistent with Google's terms of service.[4] Third, to protect the identities of the people described in this study to the extent possible, we do not plan to release our data that links GitHub users to genders.

Results

Are women's pull requests less likely to be accepted?

We hypothesized that pull requests made by women are less likely to be accepted than those made by men. Prior work on gender bias in hiring – that women tend to have resumes less favorably evaluated than men (*5*) – suggests that this hypothesis may be true.

[3] NCSU IRB number 6708.

[4] `https://sites.google.com/site/bughunteruniversity/nonvuln/discover-your-name-based-on-e-mail-address`

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

To evaluate this hypothesis, we looked at the pull status of every pull request submitted by women compared to those submitted by men. We then calculate the merge rate and corresponding confidence interval, using the Clopper-Pearson exact method (*15*), and find the following:

	Open	Closed	Merged	**Merge Rate**	95% Confidence Interval
Women	8,216	21,890	111,011	**78.6%**	[78.45%, 78.87%]
Men	150,248	591,785	2,181,517	**74.6%**	[74.56%, 74.67%]

The hypothesis is not only false, but it is in the opposite direction than expected; *women tend to have their pull requests accepted at a higher rate than men*! This difference is statistically significant ($\chi^2(df = 1, n = 3,064,667) = 1170, p < .001$). What could explain this unexpected result?

Perhaps women's high acceptance rate is because they are already well known in the projects they make pull requests in. Pull requests can be made by anyone, including both insiders (explicitly authorized owners and collaborators) and outsiders (other GitHub users). If we exclude insiders from our analysis, the women's acceptance rate (64.4%) continues to be significantly higher than men's (62.7%) ($\chi^2(df = 2, n = 2,473,190) = 492, p < .001$).

Perhaps only a few highly successful and prolific women, responsible for a substantial part of overall success, are skewing the results. To test this, we calculated the pull request acceptance rate for each woman and man with 5 or more pull requests, then found the average acceptance rate across those two groups. The results are displayed in Figure 2. We notice that women tend to have a bimodal distribution, typically being either very successful ($> 90\%$ acceptance rate) or unsuccessful ($< 10\%$). But this data tells the same story as the overall acceptance rate; women are more likely than men to have their pull requests accepted.

Why might women have a higher acceptance rate than men? In the remainder of this section, we will explore this question by evaluating several hypotheses that might explain the result.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

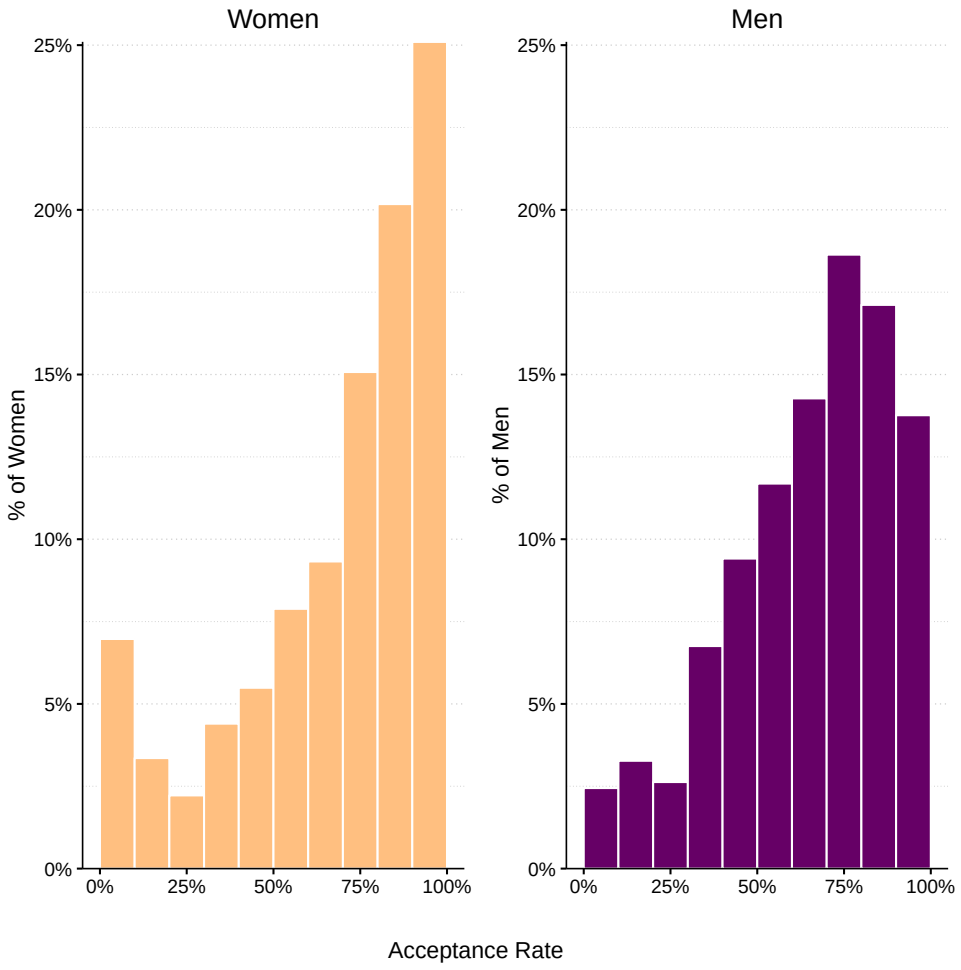

Figure 2: Histogram of mean acceptance rate per developer for Women (Mean 69.3%, Median 78.6%) and Men (Mean 66.3%, Median 70.0%)

Do women's pull request acceptance rates start low and increase over time?

One plausible explanation is that women's first few pull requests get rejected at a disproportionate rate compared to men's, so they feel dejected and do not make future pull requests. After all, Reagle's account of women's participation in virtual collaborative environments describes an unreasonably aggressive argument style necessary to justify one's own contributions, a style that many women may find to be not worth the exhaustion of proving their competence over and over (*16*). Thus, the overall higher acceptance rate for women would be due to survivorship

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

bias within GitHub; the women who remain and do the majority of pull requests would be better equipped to contribute, and defend their contributions, than men. Thus, we might expect that women have a lower acceptance rate than men for early pull requests but have an equivalent acceptance rate later.

To evaluate this hypothesis, we examine pull request acceptance rate over time, that is, the mean acceptance rate for developers on their first pull request, second pull request, and so on. Figure 3 displays the results. Orange points represent the mean acceptance rate for women, and purple points represent acceptance rates for men. Shaded regions indicate the 95% Clopper-Pearson confidence interval.

The acceptance rate of women tends to fluctuate at the right of the graph, because the acceptance rate is affected by only a few individuals. For instance, at 128 pull requests, only 103 women are represented. Intuitively, where the shaded region for women includes the corresponding data point for men, the reader can consider the data too sparse to conclude that a substantial difference exists between acceptance rates for women and men. Nonetheless, between 1 and 64 pull requests, women's higher acceptance rate remains. Thus, the evidence casts doubt on our hypothesis.

Are women making pull requests that are more needed?

Another explanation for women's pull request acceptance rate is that, perhaps, women disproportionately make contributions that projects need more urgently. What makes a contribution "needed" is difficult to assess from a third-party perspective. One way is to look at which pull requests link to issues in projects' GitHub issue trackers. If a pull request references an issue, we consider it to serve a more immediate, recognized need than an otherwise comparable one that does not. To support this argument with data, we randomly selected 30 pull request descriptions that referenced issues; in 28 cases, the reference was an attempt to fix all or part of

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

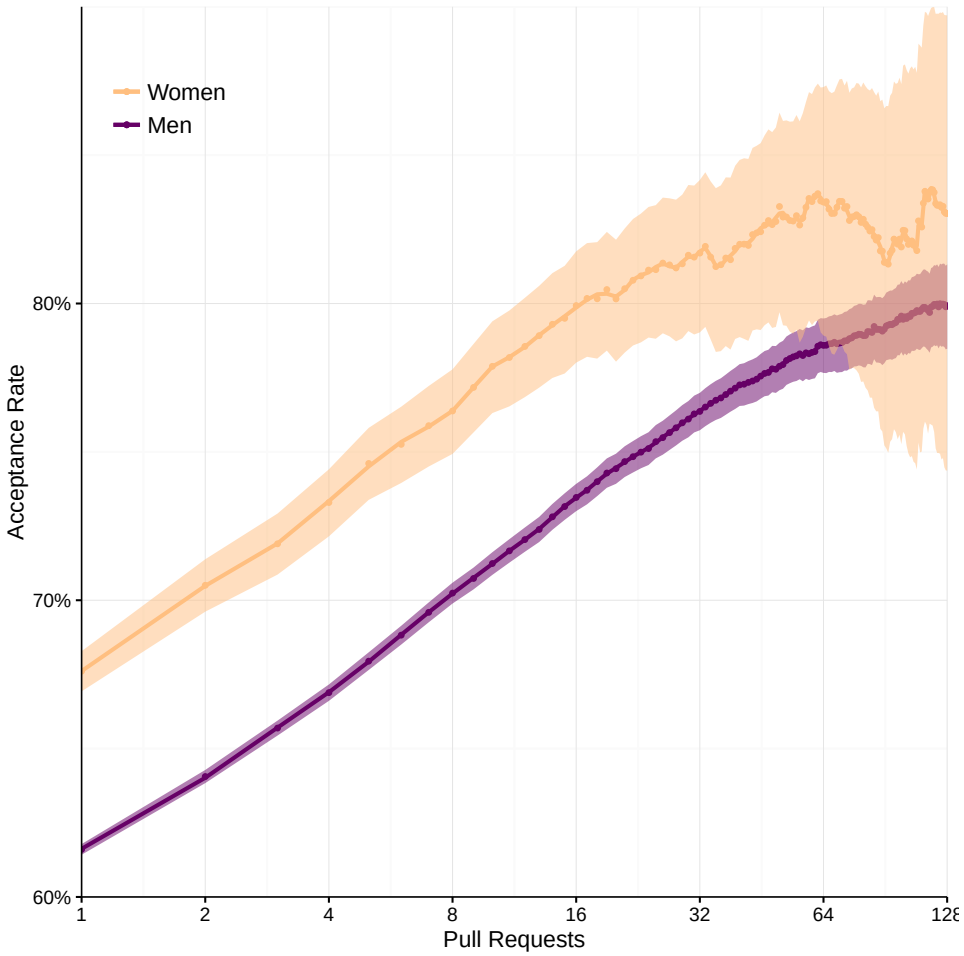

Figure 3: Pull request acceptance rate over time

an issue. Based on this high probability, we can assume that when someone references an issue in a pull request description, they usually intend to fix a real problem in the project. Thus, if women more often submit pull requests that address an immediate need and this is enough to improve acceptance rates, we would expect that these same requests are more often linked to issues.

We evaluate this hypothesis by parsing pull request descriptions and calculating the percentage of pulls that reference an issue. To eliminate projects that do not use issues or do not customarily link to them in pull requests, we analyze only pull requests in projects that have at

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

least one linked pull request. Here are the results:

	without reference	with reference	%	95% CI
Women	29,163	4748	**14.0%**	[13.63%, 14.38%]
Men	1,071,085	182,040	**14.5%**	[14.46%, 14.58%]

This data shows a statistically significant difference ($\chi^2(df = 1, n = 1,287,036) = 7.3, p < .007$). Contrary to the hypothesis, women are slightly less likely to submit a pull request that mentions an issue, suggesting that women's pull requests are less likely to fulfill an immediate need. Note that this doesn't imply women's pull requests are less valuable, but instead that the need they fulfill appears less likely to be recognized and documented before the pull request was created. Regardless, the result suggests that women's increased success rate is not explained by making more immediately needed pull requests.

Are women making smaller changes?

Maybe women are disproportionately making small changes that are accepted at a higher rate because the changes are easier for project owners to evaluate. This is supported by prior work on pull requests suggesting that smaller changes tend to be accepted more than larger ones (*17*).

We evaluated the size of the contributions by analyzing lines of code, modified files, and number of commits included. The following table lists the median and mean lines of code added, removed, files changed, and commits per pull request:

		lines added	lines removed	files changed	commits
Women	median	29	5	1	2
	mean	1591	596	5.2	29.3
Men	median	21	4	1	2
	mean	1002	431	4.9	26.8

For all four metrics of size, women's pull requests are significantly larger than men's (Wilcoxon rank-sum test, $p <= .001$).

One threat to this analysis is that lines added or removed may exaggerate the size of a change

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

whenever a refactoring is performed. For instance, if a developer moves a 1000-line class from one folder to another, even though the change may be relatively benign, the change will show up as 1000 lines added and 1000 lines removed. Although this threat is difficult to mitigate definitively, we can begin to address it by calculating the net change for each pull request as the number of added lines minus the number of removed lines. Here is the result:

		net lines changed
women	median	11
	mean	994
men	median	7
	mean	590

This difference is also statistically significant (Wilcoxon rank-sum test, $p < .001$). So even in the face of refactoring, the conclusion holds: women make pull requests that add and remove more lines of code, modify more files, and contain more commits.

Are women's pull requests more successful when contributing code?

One potential explanation for why women get their pull requests accepted more often is that the *kinds* of changes they make are different. For instance, changes to HTML could be more likely to be accepted than changes to C code, and if women are more likely to change HTML, this may explain our results. Thus, if we look only at acceptance rates of pull requests that make changes to program code, women's high acceptance rates might disappear. For this, we define program code as files that have an extension that corresponds to a Turing-complete programming language. We categorize pull requests as belonging to a single type of source code change when the majority of lines modified were to a corresponding file type. For example, if a pull request changes 10 lines in `.js` (javascript) files and 5 lines in `.html` files, we include that pull request and classify it as a `.js` change.

Figure 4 shows the results for the 10 most common programming language files (top) and the 10 most common non-programming language files (bottom). Each pair of bars represent

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

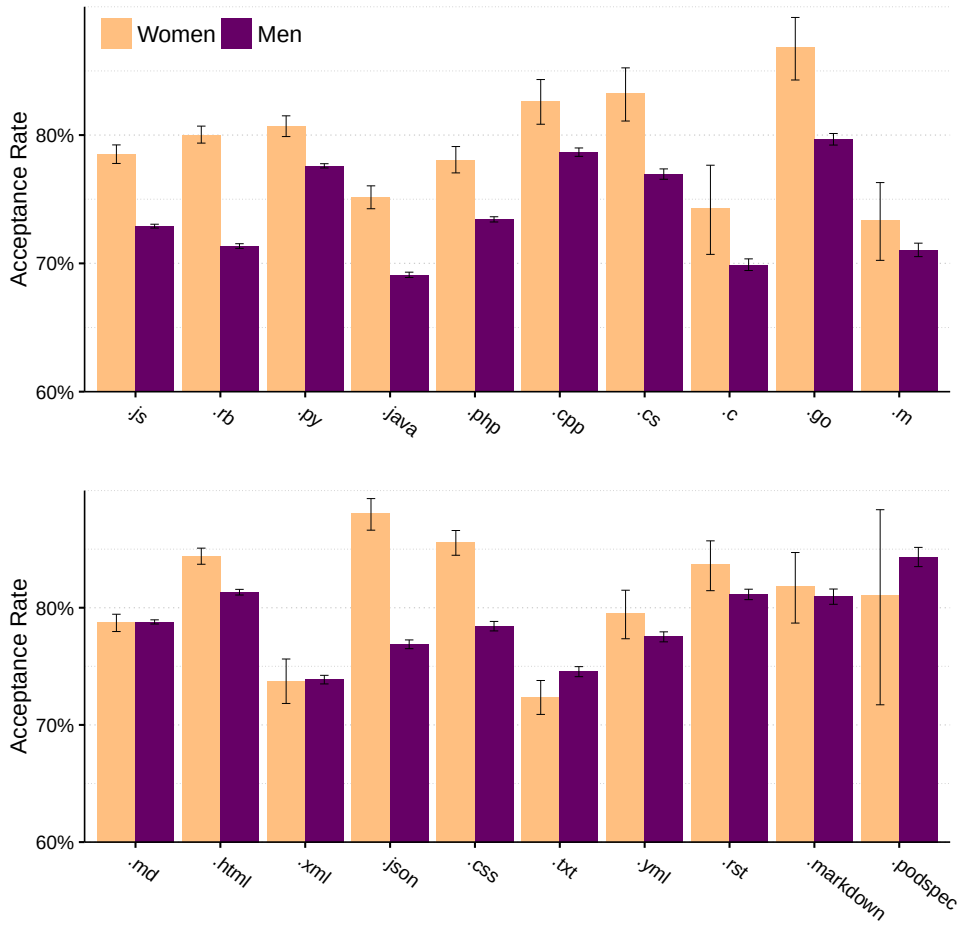

Figure 4: Pull request acceptance rate by file type, for programming languages (top) and non-programming languages (bottom)

a pull request classified as part of a programming language file extension, where the height of each bar represents the acceptance rate and each bar contains a 95% binomial confidence interval.

Overall, we observe that women's acceptance rates dominate over men's for every programming language in the top ten, to various degrees.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

Is a woman's pull request accepted more often because she appears to be a woman?

Another explanation as to why women's pull requests are accepted at a higher rate would be what McLoughlin calls Type III bias: "the singling out of women by gender with the intention to help" (*18*). In our context, project owners may be biased towards wanting to help women who submit pull requests. Thus, we expect that women who can be perceived as women are more likely to have their pull requests accepted than women whose gender cannot be easily inferred.

We evaluate this hypothesis by comparing pull request acceptance rate of developers who have gender-neutral GitHub profiles and those who have gendered GitHub profiles. We define a gender-neutral profile as one where a gender cannot be readily inferred from their profile. Figure 1 gives an example of a gender-neutral GitHub user, "akofink", who uses an *identicon*, an automatically generated graphic, and does not have a gendered name that is apparent from the login name. Likewise, we define a gendered profile as one where the gender can be readily inferred from the photo or the name. Figure 1 also gives an example of a gendered profile; the profile of "JustinAMiddleton" is gendered because it uses a login name (Justin) commonly associated with men, and because the picture depicts a person with masculine features (e.g., pronounced brow ridge (*19*)). Clicking on a user's name in pull requests reveals their profile, which may contain more information such as a user-selected display name (like "Justin Middleton").

To obtain a sample of gendered and gender-neutral profiles, we used a combination of automated and manual techniques. For gendered profiles, we included GitHub users who used a profile image rather than an identicon and that Vasilescu and colleagues' tool could confidently infer a gender from the user's name (*2*). For gender-neutral profiles, we included GitHub users that used an identicon, that Michael's tool could not infer a gender for, and that a mixed-culture panel of judges could not guess the gender for.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

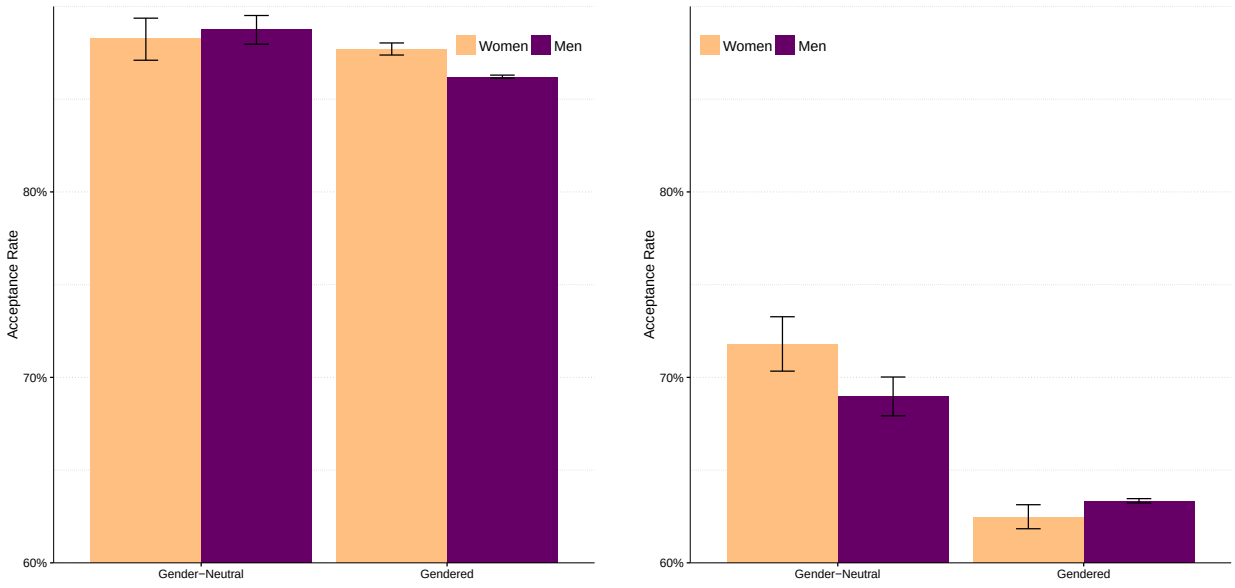

Figure 5: Pull request acceptance rate by gender and perceived gender, with 95% Clopper-Pearson confidence intervals, for insiders (left) and outsiders (right)

While acceptance rate results so far have been robust to differences between insiders (people who are owners or collaborators of a project) versus outsiders (everyone else), for this analysis, there is a substantial difference between the two, so we treat each separately. Figure 5 shows the acceptance rates for men and women when their genders are identifiable versus when they are not, with pull requests submitted by insiders on the left and pull requests submitted by outsiders on the right.

For insiders, we observe little evidence of bias when we compare women with gender-neutral profiles and women with gendered profiles, since both have about equivalent acceptance rates. This can be explained by the fact that insiders likely know each other to some degree, since they are all authorized to make changes to the project, and thus may be aware of each others' gender.

For outsiders, we see evidence for gender bias: women's acceptance rates are 71.8% when they use gender neutral profiles, but drop to 62.5% when their gender is identifiable. There is

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

a similar drop for men, but the effect is not as strong. Women have a higher acceptance rate of pull requests overall (as we reported earlier), but when they're outsiders and their gender is identifiable, they have a lower acceptance rate than men.

Discussion

To summarize this paper's observations:

1. Women are more likely to have pull requests accepted than men.
2. Women continue to have high acceptance rates as they gain experience.
3. Women's pull requests are less likely to serve an immediate project need.
4. Women's changes are larger.
5. Women's acceptance rates are higher across programming languages.
6. Women have lower acceptance rates as outsiders when they are identifiable as women.

We next consider several alternative theories that may explain these observations as a whole.

Given observations 1–5, one theory is that a bias against *men* exists, that is, a form of reverse discrimination. However, this theory runs counter to prior work (e.g., (*13*)), as well as observation 6. With 6, we observed that when a contributor's gender is identifiable, men's acceptance rates surpass women's.

Another theory is that women are taking fewer risks than men. This theory is consistent with Byrnes' meta-analysis of risk-taking studies, which generally find women are more risk-averse than men (*20*). However, this theory is not consistent with observation 4, because women tend to change more lines of code, and changing more lines of code correlates with an increased risk of introducing bugs (*21*).

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

Another theory is that women in open source are, on average, more competent than men. This theory is consistent with observations 1–5. To be consistent with observation 6, we need to explain why women's pull request acceptance rate drops when their gender is apparent. An addition to this theory that explains observation 6, and the anecdote describe in the introduction, is that discrimination against women does exist in open source.

Assuming this final theory is the best one, why might it be that women are more competent, on average? One explanation is survivorship bias: as women continue their formal and informal education in computer science, the less competent ones may change fields or otherwise drop out. Then, only more competent women remain by the time they begin to contribute to open source. In contrast, less competent men may continue. While women do switch away from STEM majors at a higher rate than men, they also have a lower drop out rate then men (*22*), so the difference between attrition rates of women and men in college appears small. Another explanation is self-selection bias: the average woman in open source may be better prepared than the average man, which is supported by the finding that women in open source are more likely to hold Master's and PhD degrees (*1*). Yet another explanation is that women are held to higher performance standards than men, an explanation supported by Gorman and Kmec's analysis of the general workforce (*23*).

In closing, as anecdotes about gender bias persist, it's imperative that we use big data to better understand the interaction between genders. While our big data study does not definitely prove that differences between gendered interactions are caused by bias among individuals, the trends observed in this paper are troubling. The frequent refrain that open source is a pure meritocracy must be reexamined.

References and Notes

1. L. Arjona-Reina, G. Robles, S. Dueas, The floss2013 free/libre/open source survey (2014).

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

2. B. Vasilescu, A. Capiluppi, A. Serebrenik, *Interacting with Computers* **26**, 488 (2014).

3. B. Vasilescu, *et al.*, *CHI Conference on Human Factors in Computing Systems*, CHI (ACM, 2015), pp. 3789–3798.

4. G. Gousios, B. Vasilescu, A. Serebrenik, A. Zaidman, *Proceedings of the 11th Working Conference on Mining Software Repositories* (ACM, 2014), pp. 384–387.

5. H. K. Davison, M. J. Burke, *Journal of Vocational Behavior* **56**, 225 (2000).

6. S. Knobloch-Westerwick, C. J. Glynn, M. Huge, *Science Communication* **35**, 603 (2013).

7. P. L. Roth, K. L. Purvis, P. Bobko, *Journal of Management* **38**, 719 (2012).

8. M. Burnett, *et al.*, *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (ACM, 2010), p. 28.

9. S. Arun, T. Arun, *Journal of International Development* **14**, 39 (2002).

10. J. W. Graham, S. A. Smith, *Economics of education review* **24**, 341 (2005).

11. V. A. Lagesen, *Science, technology & human values* **33**, 5 (2008).

12. J. Antin, R. Yee, C. Cheshire, O. Nov, *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (ACM, 2011), pp. 11–14.

13. D. Nafus, *New Media & Society* **14**, 669 (2012).

14. G. Gousios, *Proceedings of the 10th Working Conference on Mining Software Repositories*, MSR '13 (IEEE Press, Piscataway, NJ, USA, 2013), pp. 233–236.

15. C. Clopper, E. S. Pearson, *Biometrika* pp. 404–413 (1934).

16. J. Reagle, *First Monday* **18** (2012).

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

17. G. Gousios, M. Pinzger, A. v. Deursen, *Proceedings of the 36th International Conference on Software Engineering*, ICSE 2014 (ACM, New York, NY, USA, 2014), pp. 345–355.

18. L. A. McLoughlin, *Journal of Engineering Education* **94**, 373 (2005).

19. E. BrownU, D. Perrett, *Perception* **22**, 829 (1993).

20. J. P. Byrnes, D. C. Miller, W. D. Schafer, *Psychological bulletin* **125**, 367 (1999).

21. A. Mockus, D. M. Weiss, *Bell Labs Technical Journal* **5**, 169 (2000).

22. X. Chen, *National Center for Education Statistics* (2013).

23. E. H. Gorman, J. A. Kmec, *Gender & Society* **21**, 828 (2007).

Acknowledgments

Special thanks to Denae Ford for her help throughout this research project. Thanks to Jon Stallings for his help with the statistics. Thanks to the Developer Liberation Front for their reviews of this paper. For their helpful discussions, thanks to Tiffany Barnes, Margaret Burnett, Tim Chevalier, Prem Devanbu, Ciera Jaspan, Saul Jaspan, Jeff Leiter, Ben Livshits, Peter Rigby, and Bogdan Vasilescu. This material is based in part upon work supported by the National Science Foundation under grant number 1252995.

Materials and Methods

GitHub Scraping

An initial analysis of GHTorrent pull requests showed that our pull request merge rate was significantly lower than that presented in prior work on pull requests (*17*). We found a solution to the problem that calculated pull request status using a different technique, which yielded a pull request merge rate comparable to prior work. However, in a manual inspection of pull requests, we noticed that several calculated pull request statuses were different than the statuses indicated on the `github.com` website. As a consequence, we wrote a web scraping tool that automatically downloaded the pull request HTML pages, parsed them, and extracted data on status, pull request message, and comments on the pull request.

We determined whether a pull requester was an insider or an outsider during our scraping process because the data was not available in the GHTorrent dataset. We classified a user as an insider when the pull request listed the person as a member or owner, and classified them as an outsider otherwise. This analysis has inaccuracies because GitHub users can change roles from outsider to insider and vice-versa. As an example, about 6% of merged pull requests from both outsider female and male users were merged by the outsider pull-requestor themselves, which

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

isn't possible, since outsiders by definition don't have the authority to self-merge. We emailed such an outsider, who indicated that, indeed, she was an insider when she made that pull request. This problem is presently unavoidable as GitHub does not keep data on role changes.

Gender Linking

To evaluate gender bias on GitHub, we first needed to determine the genders of GitHub users.

Our technique uses several steps to determine the genders of GitHub users. First, from the GHTorrent data set, we extract the email addresses of GitHub users. Second, for each email address, we use the search engine in the Google+ social network to search for users with that email address. The search works for both Google users' email addresses (@gmail.com), as well as other email addresses (such as @ncsu.edu). Third, we parse the returned users' 'About' page to scrape their gender. Finally, we only include the genders 'Male' and 'Female' because there were relatively few other options chosen. We also automated and parallelized this process. This technique capitalizes on several properties of the Google+ social network. First, if a Google+ user signed up for the social network using an email address, the search results for that email address will return just that user, regardless of whether that email address is publicly listed or not. Second, signing up for a Google account currently *requires* you to specify a gender (though 'Other' is an option)[5], and, in our discussion, we interpret their use of 'Male' and 'Female' in gender identification (rather than sex) as corresponding to our use of the terms 'man' and 'woman'. Third, when Google+ was originally launched, gender was publicly visible by default.[6]

[5] https://accounts.google.com/SignUp

[6] http://latimesblogs.latimes.com/technology/2011/07/google-plus-users-will-soon-be-able
html

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

Merged Pull Requests

Throughout this study, we measure pull requests that are accepted by calculating developers' merge rates, that is, the number of pull requests merged divided by the sum of the number of pull requests merged, closed, and still open. We include pull requests still open in the denominator in this calculation because pull requests that are still open could be indicative of a pull requestor being ignored, which has the same practical impact as rejection.

Determining Gender Neutral from Gendered Profiles

To determine gendered profiles, we first parsed GitHub profile pages to determine whether each user was using a profile picture or an identicon. We then ran display names and login names through a gender inference program, which maps a name to a gender.[7]. We classified a GitHub profile as gendered if each of the following were true:

- a profile image (rather than an identicon) was used, and
- the gender inference tool output a gender at the highest level of confidence (that is, 'male' or 'female,' rather than 'mostly male,' 'mostly female,' or 'unknown').

To classify profiles as gender neutral, we added a manual step. Given a GitHub profile that used an identicon (thus, a gender could not be inferred from a profile photo) and a name that the gender inference tool classified as 'unknown', we manually verified that the profile picture could not be easily identified as belonging to a specific gender. We did this in two phases. In the first phase, we assembled a panel of 3 people to evaluate profiles for 10 seconds each. The panelists were of American (man), Chinese (man), and Indian (woman) origin, representative of the three most common nationalities on GitHub. We used different nationalities because

[7]This tool was builds on Vasilescu and colleagues' tool (*2*), but we removed some of Vasilescu and colleagues' heuristics to be more conservative. Our version of the tool can be found here: `https://github.com/DeveloperLiberationFront/genderComputer`

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

we wanted the panel to be able to identify, if possible, the genders of GitHub usernames with different cultural origins. In the second phase, we eliminated two inefficiencies from the first phase: (a) because the first panel estimated that for 99% of profiles, they only looked at login names and display names, we only showed this information to the second panel, and (b) because the first panel found 10 seconds was usually more time than was necessary to assess gender, we allowed panelists at the second phase to assess names at their own pace. Across both phases, panelists were instructed to signal if they could identify the gender of the GitHub profile. To estimate panelists' confidence, we considered using a threshold like "90% confident of the gender," but found that this was too ambiguous in pilot panels. Instead, we instructed panelists to signal if they would be comfortable addressing the GitHub user as 'Mister' or 'Miss' in an email, given the only thing they knew about the user was their profile. We considered a GitHub profile as gender neutral if all of the following conditions were met:

- an identicon (rather than a profile image) was used,
- the gender inference tool output a 'unknown' for the user's login name and display name, and
- none of the panelists indicated that they could identify the user's gender.

Across both panels, panelists inspected 3000 profiles of roughly equal numbers of women and men. We chose the number 3000 by doing a rough statistical power analysis using the results of the first panel to determine how many profiles panelists should inspect during the second panel to obtain statistically significant results. Of the 3000, panelists eliminated 409 profiles for which at least one panelist could infer a gender.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

Threats

One threat to this analysis is the existence of robots that interact with pull requests. For example, "Snoopy Crime Cop"[8] appears to be a robot that has made thousands of pull requests. If such robots used an email address that linked to a Google profile that listed a gender, our merge rate calculations might be skewed unduly. To check for this possibility, we examined profiles of GitHub users that we have genders for and who have made more than 1000 pull requests. The result was tens of GitHub users, none of whom appeared to be a robot. So in terms of our merge calculation, we are somewhat confident that robots are not substantially influencing the results.

Another threat is if men and women misrepresent their genders at different rates. In that case, we may have inaccurately labeled some men on GitHub as women, and vice-versa.

Another threat is inaccuracies in our assessment of whether a GitHub member's gender is identifiable. For profiles we labeled as gender-neutral, our panel may not have picked out subtle gender features in GitHub users' profiles. Moreover, project owners may have used gender signals that we did not; for example, if a pull requestor sends an email to a project owner, the owner may be able to identify the requestor's gender even though our technique could not.

Another threat is that of construct validity, whether we're measuring what we aim to measure. One example is our inclusion of "open" pull requests as a sign of rejection, in addition to the "closed" status. Rather than a sign of rejection, open pull requests may simply have not yet been decided upon. Another example is whether pull requests that do not link to issues signals that the pull request does not fulfill an immediate need.

Another threat is that of external validity; do the results generalize beyond the population studied? While we chose GitHub because it is the largest open source community, other communities such as SourceForge and BitBucket exist, along with other ways to make pull requests, such at through the git version control system directly. Moreover, while we studied a large pop-

[8] https://github.com/snoopycrimecop

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016

ulation of contributors, they represent only part of the total population of developers on GitHub, because not every developer makes their email address public, because not every email address corresponds to a Google+ profile, and because not every Google+ profile lists gender.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1733v1 | CC-BY 4.0 Open Access | rec: 9 Feb 2016, publ: 9 Feb 2016