Reasons for lack of diversity in open source: The case Katie Bouman

Lina Wang and Kristina Weinberger

February 9, 2020

1 Introduction

Open Source communities face similar problems with diversity as other areas of Computer Science. In this paper, we aim to take a look at the current situation and why diversity matters. Then, we discuss the relatively recent events around Katie Bouman and the first picture of a Black Hole as an example for some of the possible reasons for a lack of women. This is followed by an overview of other common issues that may contribute to imbalances in Open Source contributions. Finally, we spend some time looking at positive developments and initiatives aimed at bringing in more diversity.

1.1 Current Status of Diversity in Open Source

It is well known that women and other minority groups tend to be underrepresented in computer science in general. This lack of diversity also extends into open source projects. According to a survey among more than 5000 randomly selected github[1] users, more than 90% of all contributors to open source projects identify as male, while only around 3% identify as female [24]. The survey results show a lower percentage of women, non-binary people and racial minorities than present in the wider area of computer science [7], which is already well below the general population.

While this survey only shows a narrow slice of the entire open source community and other communities around free and open technologies, it is reasonable to expect a relative lack of women and other minorities overall. Research and discussion on diversity in open source often focus on women specifically, which is also true for this paper. Other minorities are often less visible. Many surveys include questions regarding ethnicity, gender or sexual orientation, but other areas, like disability or social class are rarely mentioned. Factors not included in most surveys do not necessarily imply under-representation. There are indications that there is a relatively high percentage of LGBTQ+ people active in open source communities [7], even though this is not included in all surveys.

1.2 Benefits of Diversity

One of the more popular theories explaining the lack of diversity in communities is that women or other underrepresented groups are less suited to the field in question. However, even if this would be true, there are still benefits to be gained from forming diverse teams.

As multiple studies [13, 21, 22] have shown, more diverse groups perform better across several metrics, such as creativity, product quality, inclusivity and making all voices heard. There is also a proven link between company staff diversity and financial earnings, suggesting that there is a lot of value in diversity. Those studies generally focus on different

[1] https://github.com

industries and for-profit companies, but the effects shown may also contribute to a higher code quality in more diverse open source communities.

Another benefit of diverse groups in software development specifically is the inclusion of more perspectives, which generally result in a product that is usable for a wider audience. Various past faux pas could have been avoided by having more inclusive teams, such as voice assistants that do not listen to women [3] or image recognition algorithms that do not recognize people of colour [19].

In recent years, advancements towards more diversity have been made and its benefits have been sufficiently documented to convince many of the necessity of fostering diverse environments. However, as the statistics mentioned above show, communities either fail to grow to be diverse on their own or do so very slowly. This paper aims to illustrate at least some of the hurdles that prevent this and show possible ways to work towards more inclusivity.

1.3 Motivation

As computer scientists and women, diversity in Open Source and Computer Science, in general, is relevant to us personally, which sparked our initial interest in this area. The lectures in Free and Open Technologies, along with the fact that almost exclusively white men were featured as pioneers of Open Source, led to the question why there were no visible individuals of other genders or ethnicities. In this text, we take a look at some potential reasons for a lack of women in open source software communities specifically. However, many of the reasons listed affect many minorities in different areas as they coincide with general societal issues and biases. As a starting point, we chose to take a closer look at the reactions around Katie Bouman presenting the first picture of a black hole in 2019. While her story is unfortunately not unique, it provides an example for some of the general issues women in open source projects face.

2 The case: Katie Bouman

In order to show why open source communities have a diversity problem, we analyzed a recent event in this area: the first-ever picture of a black hole. From this event, we took a closer look at the reactions happening because of one woman scientist - Katie Bouman, who helped capture the first image of the black hole. This section should provide a glance behind the historical moment and point out diversity issues concerning open source communities.

2.1 Background story

The first visualization of the black hole was unveiled in April 2019. After this celebratory image was published, another picture showing the reaction of a young female computer scientist, named Katie Bouman, went viral. Bouman, a postdoctoral fellow at MIT and a member of the team running Event Horizon Telescope, made a considerable contribution with her algorithm to capture this image. There was a team of around 200 researchers who contributed to the breakthrough. However, media quickly began hailing Bouman as the face of the black hole project and turned her into a “lone genius” for women in STEM. In fact, Bouman tried to ensure that there would not be one person singled out over any other:

> “No one algorithm or person made this image. It required the amazing talent of a team of scientists from around the globe.”

On the one hand, Bouman stood out as a role model as she is a woman who works in a men-dominated field. On the other hand, it ended up on receiving end of misogynistic trolling and sexist speculations because she is a woman in STEM. Her Wikipedia page was even proposed for deletion. Furthermore, she was accused of taking credit for her male colleague's work, Andrew Chael, because he had made 850.000 commits to the GitHub repository while Katie Bouman made only 2.400. Although Chael did write a good amount of code, most of those commits are models and data.[20]

2.2 Relation to Open Source

Free and Open Source software was at the heart of this image. The use used different imaging software libraries to achieve the goal. Out of the three, two were fully open source libraries - `eht-imaging` and `SMILI`. Both libraries are available on GitHub under GNU GPL v3. [1]

The reactions happening around Katie Bouman's case give a reasonable explanation for the lack of diversity in open source communities. The following section should further elaborate on these issues derived from Bouman's incidence and analyze how these issues relate to open source in general. The three issues we looked into are meritocracy in open source, lone genius, and gender bias on Wikipedia.

2.3 Meritocracy

Michael Young primarily coined the term meritocracy in his book in 1958. Young used this term, satirically that intelligence and merit are favored above everything else. Meritocracy is now presented as a social ideal which commits to diversity. Diversity applies not only to gender but also to underrepresented minorities. Open Source communities seek for this concept as well. In the open-source community, meritocracy describes that everyone, regardless of social status, gender, race, religion has access to the same information, and only the quality of contribution matters. In software projects, contribution usually means writing the code. Those who write most of the code, have the most merit and are most deserving.

2.3.1 Risks and Effects in Open Source

Nevertheless, meritocracy has been criticized. The concept has become more linked to hidden bias and outright abuse. The preference involves devaluing person contributions based on specific aspects such as gender, race, and so forth. As a result of meritocracy, people who are underrepresented in STEM will be excluded. Because it benefits the privileged one - not only the one with skill and ability but with self-confidence to demonstrate expertise publicly. Especially in the open-source culture, it disadvantages people who cannot afford to spend their time giving away their work for free.

In fact, emphasizing the value of merit can actually lead to more bias in favor men. In Castilla's study [4], she analyzed one company that had implemented a merit-driven compensation system. But results have shown that salary increases for women with the same quality of contribution were significantly lower - as well as for ethnic minorities, and none US-born employes. Another study [25] of gender bias in open source found that women who contributed openly got their pull requests to merge less often than when they hid their gender.

Referring to the case of Katie Bouman, people were trying to discredit her work on the black hole project, because Andrew Chael wrote most of the lines. This reaction shows the once again the meritocratic myth. Those who write more code is merit. Nevertheless, people can add many lines of code to a project that adds relatively little value. Commits

to Github hardly correlates with success as an engineer. A study [27] has also shown that GitHub commits measurement is an estimate that is usually wrong and counts many things that are not actually code. Besides, writing code is not the only contribution that can be considered. There are also plenty of non-code contributions, which would include organizational tasks. Women will often conduct the non-code part of work.

Another example, where GitHub commits were taken as a measurement for the quality of contribution, was Rafael Avila de Espindola. He was one of the contributors to the Low-Level Virtual Machine project (LLVM). When he left the open-source project, he gave the Code of Conduct as a reason for his leaving. In this article [18] Espindola was presented as one of the most worthy contributors because of the large number of commits he has.

2.4 Putting individuals in the spotlight

Looking back at history, there are examples for "lone genius" in STEM, in which one person is responsible for an ingenious creation. Sir Isaac Newton, for example, is credited with much of not only the physics but also the mathematical framework. Although there were other scientists, who were also working on the development of calculus, Newton is presented as the lone scientist hero. Johnson et al. [11] made the point that breakthroughs come from "group genius" and not lone epiphanies.

The image of "lone genius" has been questioned by Gingras et al. [14], because science is far from a lonely endeavor. Especially in the open-source community, developers contribute to a project because of common interests. Open source projects gain success through collaborative work and a combination of different skill sets.

In the case of Katie Bouman, seeking to credit one person for an open-source project is massively simplistic from a technical perspective, and it is ambiguous to what open source aims to. The black hole image project was a result of collaborative work. There was a huge team behind the black hole image. Bouman made an outstanding contribution, but she was just one of many.

2.4.1 Effects

Putting Bouman in the spotlight of the whole black hole image event, she was judged by the public because she is a woman. Issues will arise as soon as the scientific results are released to the public. There are people around the internet who have the misperception that Bouman is just a minor contributor to a project where men like Andrew Chael deserve the credit. Studies [17][23] have exposed a direct link between the expectation of "genius" for discipline and the gender and race gap for negatively stereotype minorities. People tend to associate the adjective "genius" to males far more frequently than to female scientists.

2.5 Male dominance on Wikipedia

Wikipedia is a web-based editing platform where anyone edits and openly collaborates. It is indeed inspired by the movement of free and open-source software. However, there is a gender bias in Wikipedia biographies. The existing bias can be proven with the case of the science professor Donna Strickland. She was approved for a Wikipedia page only after winning the Nobel Prize for Physics in 2018. Before that, Wikipedia deemed her unworthy of recognition. [2]

The cases of Bouman and Strickland are representative examples of gender inequity issues that Wikipedia deals with.

3 Other Issues

As we discussed, the reactions to Katie Bouman show some of the reasons why open source communities may lack in diversity. This section discusses several wider cultural issues that may be present in a community. These are usually not enforced on purpose, but a by-product of communities growing in a field lacking diversity and general societal conventions.

3.1 Toxic Environments

Many open source communities have a reputation of being a very harsh environment. There are many stories from people who got their questions and contributions met with derision and ridicule, which does not create a welcoming environment[2]. Linus Torvalds has long been infamous for his scathing remarks towards people trying to add their own additions [15], but his projects are not the only example of this by far. The expectation and the witnessing of negative interactions is one of the main reasons people gave for not contributing more in Github's open-source survey [24] or stopping to contribute to a specific project.

Several people have left open source communities [10, 8] because they did not feel comfortable in the toxic environments created by a lack of empathy for other people's feelings. Breaking into a community that acts hostile towards both members and outsiders requires a large amount of motivation and commitment. This tends to hamper people who do not fit the mold of the traditional white, male computer nerd more, as they have usually been exposed to more questioning about their abilities and motivations, while also often having fewer role models and less support.

While this environment affects all people trying to be active in a community, many forms of negative comments draw on general stereotypes about marginalized groups. In addition, harassment of minorities will often be more tolerated and less visible as inappropriate behaviour. Complaints about targeted abuse may be ignored, which makes a space even less hospitable for the people this is aimed at.

The tendency towards putting certain people in the spotlight may exacerbate this issue, especially when the person in question is one of the main harassers or abusers in the community. Due to the perceived value of that person's contributions, accusations will often be seen as less convincing.

3.2 Open Sexism

Arguably, Bouman's experiences can be seen as an example for open sexism, as similar stories featuring men as figureheads did not garner similar reactions. Although the projects credited to men are often just as much a group effort as Bouman's black hole project, naming them as a single person's achievement is rarely questioned. It's also notable that a lot of the reaction turned against Bouman as a person, despite her consistently pointing out the group behind her. On the other hand, it's possible to assume, giving the benefit of doubt, that all those reacting to Bouman being credited would have reacted similarily to the media picking Chael as the face accompanying the black hole picture.

In addition to potentially more negative reactions towards women standing out, there are many examples of sexist culture existing in open source communities. Code features jokes or variable names objectifying and disparaging women [9, 6] and sexist remarks may be acceptable behaviour on a public mailing list. Attempts to change this culture

[2] See for example https://news.slashdot.org/story/14/10/06/1837237/lennart-poettering-open-source-community-quite-a-sick-place-to-be-in, as the original article is no longer available

or to remove unnecessarily offensive comments have often been met with considerable resistance [5].

Many communities, especially larger ones, regularly have conferences and meetups focused around their open source project. At those events, as well as other male-dominated events, women's experienced tend to be impacted by the presence of inappropriate jokes, others assuming they do not belong or sexual harassment. While this is an issue more conferences are starting to take seriously, there have been many complaints that did not result in permanent change. However, recent attention to this problem has led to an increasing number of conferences adopting Codes of Conduct and establishing procedures for handling incidents. Those developments are discussed in more detail below.

3.3 Existing Biases

In addition to the more overt issues discussed above, there are also more subtle and unconscious biases against women's contributions. These biases are generally not the result of intentions or conscious thoughts but have been repeatedly shown in various studies.

For open-source projects specifically, a study was done by Terrell et al. [25], has shown that pull requests on Github are less likely to be accepted if they are opened by women. In this study, they found no evidence for a lower quality of code in women's contributions. This further lowers the amount of diversity in open source projects, as female developers face a higher barrier towards successful contribution.

Similar effects have also been shown in other fields, for example, in musical auditions, paper acceptance processes at conferences [12], or code reviews [16]. If there was no indication of whether the submission was by a man or a woman, acceptance rates for women were significantly higher. Further studies on the effects of sexist and racist biases on hiring processes have also shown that white men are more likely to be perceived as a good fit for a position, even if there is no difference between listed qualifications.

This tendency to devalue the achievements of women has also affected the documentation of history. While female scientists are now facing fewer hurdles in getting their results accepted as new knowledge, their names are rarely associated with their achievements. Aside from denying recognition to the scientists themselves, this also leads to a lack of role models for future generations. From this perspective, portraying Katie Bouman as the person responsible for successfully creating a picture of a black hole is an important step towards more diversity.

4 Positive Developments

In recent years, the open-source community operates on a set of values, including a commitment to inclusion and diversity. The following section will elaborate on a few positive developments, such as a code of conduct and double-blind reviews.

4.1 Code of conduct

One of the ways open source projects attempt to push diversity is including a Code of Conduct (CoC) [26]. The aim of establishing a CoC is to support communication among developers and prevent conflicts during the work process. It sets a ground rule for communication between members, so everyone contributing to open source projects can feel comfortable. A CoC should reduce explicit gender-based biases and discrimination towards women in STEM. In September 2018, after Linus Torvalds stepped back, they introduced their CoC, though controversial.

The Contributor Covenant is a code of conduct for open-source projects. Many widely used and influential platforms, such as Creative Commons, Linux, and even Google, utilize the Contributor Covenant. However, a CoC does not protect people from harassment or discrimination. There must be specific roles such as project maintainer who needs to make sure that the guidelines are followed. They are expected to be committed to enforcing the CoC, including dealing with problems as they emerge.

4.2 Double-Blind Reviews

It is known that the scientific peer-review process is also not free from bias [12]. Male editors choose more males and females more females. Kaatz et al. study also show that women face a disadvantage in general because there are relatively fewer women in the process of scientific publication. As transparency is one of the fundamental pillars in open science, double-blind reviews conceal identities of both authors and reviewers. One of the benefits of double-blind reviews is the possibility of reviewer bias, which can take many forms, such as an author's gender, country of origin, and academic status.

Mozilla, for example, is making incremental progress in narrowing the gender divide. They want to make the code review process more egalitarian, especially in open source projects. The game developers a way to anonymize pull requests, so reviewers would see the code but not the identity of the person who wrote it. When it comes to diversity and inclusion, Mozilla is a leader in the open-source community. [16]

4.3 Women in Red

The Women in Red (WiR) project aims to tackle the gender bias posed by Wikipedia. It was launched to feature more notable women throughout history on Wikipedia proactively. One of the tools is called an edit-a-thon. WiR and universities worldwide promote these events in which people gather together to edit Wikipedia biographies about women and related issues.

5 Conclusion

The fact that women are underrepresented in tech is nothing new, but the numbers for women in open-source projects are even worse. A survey on Github users could confirm this finding.

In this work, we take the case of Katie Bouman in attempting to explain why open source communities are lacking diversity. In looking at the reactions happening around Bouman's case, several issues related to open-source communities could be identified. One of the common issues is the meritocratic ideal, where its mission is to push diversity. However, we found that the concept produces a contrary effect. It raises inequality than diversity instead. Another problem derived from Bouman's case is the effect of "lone genius". Through our research, we can conclude that this idea of making one individual responsible for a breakthrough should not be pursued. Instead of this idea, we should be moving toward an image of "group geniuses". Furthermore, we found several wider cultural issues in open-source projects in general.

According to our findings of positive developments, there are several commitments to inclusion and diversity. The introduction of Codes of Conduct into many projects is one example for efforts to reduce gender biases and discrimination towards women in the open-source community. Another potential solution to this perceived gender bias is for code repositories to adopt double-blind reviews where neither the authors nor the reviewers are aware of each other's identity and gender. Finally, a gender gap bridging project on Wikipedia was launched to fix the disparity in articles about noteworthy men and

women, giving more recognition to deserving women and presenting role models to future generations. While the issues we have described still remain relevant today, we feel that open source is actively working towards more diverse communities.

References

[1] Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *The Astrophysical Journal Letters*, 875(1):L4, 2019.

[2] Salvatore Babones. Why did donna strickland not have a wikipedia page? *published on https://nationalinterest.org/feature/why-did-donna-strickland-not-have-wikipedia-page-34702*, 2018.

[3] Joan Palmiter Bajorek. Voice recognition still has significant race and gender biases. *published on https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases*, 2019.

[4] Emilio J Castilla and Stephen Benard. The paradox of meritocracy in organizations. *Administrative Science Quarterly*, 55(4):543–676, 2010.

[5] Rosalie Chan. A group of women trying to change the sexist culture of open source software have been harassed online. *published on https://www.businessinsider.com/women-running-for-the-open-source-initiative-face-online-harassment-2019-3?r=DE&IR=T*, 2019.

[6] Jonathan Corbet. Who controls glibc? *published on https://lwn.net/Articles/753646/*, 2018.

[7] Klint Finley. Diversity in open source is even worse than in tech overall. *published on https://www.wired.com/2017/06/diversity-open-source-even-worse-tech-overall/*, Feb 2017.

[8] GNOME Foundation. In memory of telsa gwynne. *published on https://www.gnome.org/news/2015/11/in-memory-of-telsa-gwynne/*, 2015.

[9] Matthew Garrett. Microsoft's ill-chosen magic constants. *published on https://mjg59.dreamwidth.org/14955.html*, 2012.

[10] Jon Gold. Linux community acts after years of complaints like sarah sharp's. *published on https://www.networkworld.com/article/2988850/linux-kernel-dev-sarah-sharp-quits-citing-brutal-communications-style.html*, 2018.

[11] Steven Johnson and Where Good Ideas Come From. The natural history of innovation, 2010.

[12] Anna Kaatz, Belinda Gutierrez, and Molly Carnes. Threats to objectivity in peer review: The case of gender. *Trends in pharmacological sciences*, 35:371–373, 08 2014.

[13] Eric Kearney, Diether Gebert, and Sven C Voelpel. When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management journal*, 52(3):581–598, 2009.

[14] Vincent Lariviere, Yves Gingras, Cassidy R. Sugimoto, and Andrew Tsou. Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332, 2015.

[15] Sam Machkovech. Linus torvalds on why he isn't nice: "i don't care about you". *published on https://arstechnica.com/information-technology/2015/01/linus-torvalds-on-why-he-isnt-nice-i-dont-care-about-you/*, 2015.

[16] Judy McConnell. Mozilla experiment aims to reduce bias in code reviews, Mar 2018.

[17] Meredith Meyer, Andrei Cimpian, and Sarah-Jane Leslie. Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in psychology*, 6:235, 2015.

[18] Larabel Michael. One of llvm's top contributors quits development over coc, outreach program. *published on https://www.phoronix.com/scan.php?page=news_item&px=LLVM-Rafael-Espindola*, May 2018.

[19] Jethro Mullen. Google rushes to fix software that tagged photo with racial slur. *published on https://edition.cnn.com/2015/07/02/tech/google-image-recognition-gorillas-tag/index.html*, 2015.

[20] Brian Resnick. Male scientists are often cast as lone geniuses. here's what happened when a woman was., Apr 2019.

[21] Miki Tsusaka Matt Krentz Rocío Lorenzo, Nicole Voigt and Katie Abouzahr. How diverse leadership teams boost innovation. *published on https://www.bcg.com/en-us/publications/2018/how-diverse-leadership-teams-boost-innovation.aspx*, 2018.

[22] Courtney Seiter. The 3 research-backed benefits of diversity that guide our team growth. *published on https://open.buffer.com/diversity-benefits/*, 2017.

[23] Daniel Storage, Zachary Horne, Andrei Cimpian, and Sarah-Jane Leslie. The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and african americans across fields. *PLOS ONE*, 11(3):1–17, 03 2016.

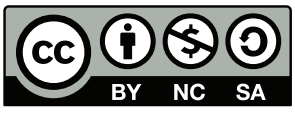
[24] Github Open Source Survey. https://opensourcesurvey.org/2017/.

[25] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson R Murphy-Hill, and Chris Parnin. Gender bias in open source: Pull request acceptance of women versus men. *PeerJ PrePrints*, 4:e1733, 2016.

[26] P. Tourani, B. Adams, and A. Serebrenik. Code of conduct in open source projects. In *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 24–33, Feb 2017.

[27] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark G.J. van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. Gender and tenure diversity in github teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 3789–3798, New York, NY, USA, 2015. Association for Computing Machinery.

About this document. This seminar paper was written as part of the lecture *Free and Open Technologies*, held by Christoph Derndorfer and Lukas F. Lang at TU Wien, Austria, during the winter term 2019/2020. All selected papers can be found online.[3]



[3] https://free-and-open-technologies.github.io