# Gender and Participation in Open Source Software Development

HANA FRLUCKAJ, LAURA DABBISH, DAVID WIDDER, SOPHIE HUILIAN QIU, and JAMES HERBSLEB*, Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University

Open source software represents an important form of digital infrastructure as well as a pathway to technical careers for many developers, but women are drastically underrepresented in this setting. Although there is a good body of literature on open source participation, there is very little understanding of the participation trajectories and contribution experiences of women developers, and how they compare to those of men developers, in open source software projects. In this study, we conducted a set of interviews with 23 developers, 11 men and 12 women, who became core in an open source project in order to understand their joining and participation trajectories. We identify differences in women and men's motivations for initial contributions and sustained involvement in a project. We also describe unique negative experiences faced by women contributors in this setting in each stage of participation. Our results have implications for diversifying participation in open source software and understanding open source as a pathway to technical careers.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: open source software, gender, diversity, inclusion, open collaboration

**ACM Reference Format:**
Hana Frluckaj, Laura Dabbish, David Widder, Sophie Huilian Qiu, and James Herbsleb. 2022. Gender and Participation in Open Source Software Development. In *Under Review - CSCW '22: ACM Conference on Computer-Supported Cooperative Work, Month Dates, 2022, Location.* ACM, New York, NY, USA, 30 pages. https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

Despite a culture widely advertised as "meritocratic", the representation of women in open source software remains notoriously low. For example, in GitHub's 2017 large-scale open source survey, only three percent of respondents identified as women [39]. At the same time, there is extensive evidence for the value in diversity— gender and racial diversity are associated with greater creativity and performance in teams [47], greater earnings (sales, revenue, profit) in organizations [78], and higher productivity in open-source projects [106]. Enhancing diversity of open source software projects should enhance their performance, improving their efficiency and the reliability and innovativeness of the code they produce.

Participation in open source software is also increasingly important for obtaining employment in technical fields. Finding ways to make open source communities more welcoming to women could potentially nearly double the available workforce, with major benefits both to industry, where


*Under Review - CSCW '22, Month Dates, 2022, Location*

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

technical talent is in short supply, and to society by greatly enhancing our ability to maintain open source software as a digital resource.

Recent work has identified specific barriers to making initial contributions to open-source (e.g., [89, 91]), and systematic biases by authority holders, e.g., imbalances in pull request acceptance rates [97]. It is relatively unknown, however, how underrepresented newcomers, particularly women, become integrated as full members into open-source communities, and why this happens so rarely. Although previous research has examined the joining process it has largely been at the project level rather than focused on individuals' participation experiences. Research on motivations for participating in open source focuses on the contributor's perspective, but we lack a detailed understanding of how motivations evolve over the course of joining an open source project in the face of positive and negative experiences. And only a few studies on OSS motivations have even considered women contributors. We lack an in-depth understanding of what motivates women versus men contributors to initially engage with an open source project and increase their involvement over time.

In this study we aim to better understand how the pathway to embedded contributor status differs for women versus men in terms of their motivations for participation over time. We also seek to understand the unique experiences faced over the different stages of the joining process for women in an environment where they are so underrepresented. Specifically, we examine the following research questions:

- *(RQ1) What motivates women developers, as compared with men, to initially engage with an open source project?*
- *(RQ2) What motivates women developers, as compared with men, to increase their involvement over time?*
- *(RQ3) What unique experiences do women face at each stage of participation in an open source project?*

In order to address these questions, we conducted a qualitative interview study with 11 men and 12 women in the PyPi open source ecosystem. We recruited developers who were actively involved in an open source project on GitHub, focusing in our interviews on their history of participation in the project and key experiences that drew them in or pushed them away. We found that both women and men made initial contributions to a project out of technical need and were assigned to do so as part of their work or out of personal interest. Once more involved in a project, both men and women engaged in broad forms of work, including contributions to social infrastructure through community building and project advocacy. We found a set of negative experiences led to decreased participation on a project or dropout, including conflict over code decisions and stress from working with entitled users. We consider unique differences in participation experience for women contributors, finding that women deal with the pressure of being highly visible, gender non-inclusivity, excessive scrutiny, and a fear of looking stupid in an environment that is over 95% men. Our results have implications for diversifying participation in open source software and understanding open source as a pathway to technical careers.

## 2 RELATED WORK

### 2.1 Overview: Open Source Software

Open Source Software (OSS), at its most basic, is code licensed such that anyone can freely inspect, modify, or re-purpose it, with minimal restriction [61]. However, the term usually also connotes the social dynamics with which the code is produced: often as a collective of distributed volunteers working together [34]. Projects may have a small set of people doing the majority of the new code development [70], who are often said to have reached the 'core' of the 'onion', with those in

outer layers contributing small patches or reporting bugs [19]. However, more recent work has suggested that the onion model may be outdated, given that modern open source projects often exist as part of an interrelated ecosystem instead of a large standalone project [53]. Jergensen et al. found that across six projects in the GNOME ecosystem (a large, well-established F/LOSS project), only about 23.3% of contributors followed some portion of the onion model. This departure from the Onion model is corroborated by findings from Herriaz et al., which suggest that volunteers and hired developers exhibit different joining patterns. While the volunteer subgroups follow the onion model, more advanced or trusted developers are quickly granted central access to the project. [45]

Open source software is often produced on social coding sites such as GitHub, which not only serves as a platform to host code, but makes information about users and their activities transparently visible, allowing users to manage their reputation, coordinate their work, and build their skills [22]. In this section, we discuss related work in the literature on open source software, notably barriers to newcomers (section 2.2), literature on later stages of joining (section 2.3), concluding with literature on gender issues in open source development and how these affect this joining process (section 2.5).

### 2.2 Newcomer Barriers in Open Source Software

The barriers newcomers to open source software projects face is an issue of intense study. Steinmacher et al. used a literature review, feedback from students attempting to contribute to open source, questionnaires, and semi-structured interviews to examine barriers newcomers face while making their first open source contribution [90, 92]. The authors grouped the 58 barriers they found into six categories: cultural differences, newcomers' characteristics, reception issues, orientation, technical hurdles, and documentation problems [89, 90]. They noted that some of the barriers they found are similar to other online production environments like Wikipedia [90, 116]. The authors later develop a portal to help newcomers make their first contribution, and show using qualitative and experimental methods that it helped newcomers find a task to start with, increased their self-efficacy, and sometimes helped facilitate interaction with the community, among other benefits [91]. Other newcomer support tools include features that classify easy first issues for newcomers [58, 88], automatic recommendations of mentors [13, 14] and of code reviewers [115], tool design approaches that maximize gender inclusiveness [66], visualization support to assist newcomers in navigating a project's issue tracker [108], and finally a tool to store a project's institutional memory to help newcomers begin work faster [20].

Evidence on the effect of mentorship on newcomer onboarding is mixed: one study showed that mentoring helps newcomers become active more quickly, and that more mature projects are best equipped to provide mentorship [28], but another study on a different open source project found that newcomers participating in a mentorship program were less likely to become long term contributors [58]. Mentors may also face barriers: some mentors find it difficult to choose newcomer-friendly projects, and struggle to identify appropriate tasks for newcomers [4].

### 2.3 Beyond the Newcomer: Later Stages in Open Source Contribution

What are the experiences of those who successfully overcome newcomer barriers, and become more involved contributors? This later process is variously referred to in the literature as developer initiation [41, 86], entering the circle of trust [86], migration [52], or immigration [9]. Being granted *committer* status, that is, the ability to directly modify ("commit") source code and integrate the contributions of others, is widely seen as a significant event in this process [9, 15, 25, 41, 52, 107], and only those contributors who have sufficiently proven themselves through their activities become committers [31, 55, 79]. One mixed method study shows that when core contributors leave a project, new contributors can take over to ensure the project's survival, but when new

contributors were surveyed, they cited not being given commit access as a barrier to their continued contribution [1]. Quantitative studies show that factors such as submitting new code to demonstrate competence [9, 86], being centrally located in the project's social network [9, 41], or working for an organization which sponsors the project [86] are associated with an increase in advancing to *committer* status.

There are several case studies of the socialization and joining processes in established open source communities: one study compared three such communities using interviews and quantitative modeling, finding that advancement is dependent on a record of collaborative technical contribution, as well as other rites of passage according to the tradition of specific communities [52]. A case study of a different community used methods including interviews and quantitative code repository and email analysis to detail the advancement process in that community [107]. They introduced the notion of community-defined "joining scripts" newcomers had to follow to become members. In their case this joining script consisted of demonstrating the appropriate level and type of activity in order to be granted access to the community version control system. Finally, another study used short surveys of recently disengaged contributors and quantitative modeling of their contribution history, and found that job or other life changes usually preceded their wholesale dropout from open source contribution [68].

However, many of these studies focus on ecosystem-level factors that are associated with advancement, rather than the specific experience of the individuals involved. Additionally, most of the literature on the joining and socialization process in open source centers the project as the unit of analysis, rather than centering the developer as the unit of analysis, whose unique goals, background, and specific experience affect this process and their future participation. Additionally, using the project rather than participants as a unit of analysis precludes analysis of the experiences of developers who contribute to multiple projects simultaneously or switch projects in response to changing life or career circumstances. We believe this gap in the literature presents an opportunity to center experiences which catalyze or dissuade experienced developers' involvement in open source communities, much in the same way past literature has done this for newcomers.

At the participant level, there is a substantial body of literature on developer motivations in open source software (e.g., [38], [79], [59],[29], [99]). This work finds that contributors are motivated by a mixture of intrinsic and extrinsic factors, including social capital, identifying with the community and feeling obligated to contribute back [46]. Much of the research on motivation, however, was conducted more than 10 years ago, before the introduction of the GitHub platform and heavy commercial involvement. In addition, this work included only male developers as participants or did not specify gender. More recently, Gerosa et al. surveyed 242 open source contributors on their motivations for contribution across their career, finding that social and reputational motivations were more prevalent than in past work, that intrinsic motivations remained important and that motivations varied for early versus longer term contributors [102]. Gerosa and much of the research on developer motivation focused at the career, rather than project level. We believe there is an opportunity to delve more deeply into the process of becoming embedded on a specific project and center experiences which catalyze or dissuade embedded developers' involvement in specific open source communities much in the same way past literature has done this for newcomers. Additionally, as noted by [100] in their survey of the literature on women's participation in open source, we lack an in-depth analysis of women's motivations in open source software development. In order to address this gap, in the current study we contrast project-level participation motivations and experiences for women versus men.

### 2.4 The Importance of Social Factors

Research on self-organizing networks suggests that people tend to prefer (i) repeated collaborations (cohesion) over new ones, to benefit from prior interactions, greater mutual trust, and increased knowledge about each other's technical and social skills, and (ii) collaborations with established (higher status) actors, which increases their own motivation and their perception of the likelihood of project success [42, 43]. Teams composed of members with prior joint experience can coordinate their efforts and leverage each other's expertise more effectively [30, 63]. Because of this prior joint experience, over time, the teams will engage in more meaningful interactions, adjust to the surface-level differences between them, and benefit from their different cognitive frameworks and value sets (i.e., their information processing [81] approaches), thus improving the team's efficiency and decision-making processes [44].

In open-source software development, contributors are motivated by a mixture of intrinsic and extrinsic factors. Aspects related to social capital, such as identifying with the community and feeling obligated to contribute back, are highly important [46]. There is a rich body of empirical research on open-source showing how identification, obligation, emotional attachment, trust relationships, and shared goals and norms, all of which are more likely to develop in cohesive project teams [113], positively impact individual and team outcomes. Prior work has found, e.g., that open-source developers who form more cohesive ties with other contributors, have a greater sense of social identity, or report higher levels of identification and obligation towards their respective communities tend to be more productive [3, 96, 113] and experience greater trust [93]. It follows that social capital should positively impact the contributors' level of engagement with an open-source project. Typically in open-source software, participants are free to disengage at any time. Therefore the extent to which they have a sense of social identity, or perceive themselves to be part of the community, may substantially increase their intention to continue and expand their level of commitment and effort [72, 87].

### 2.5 Gender Representation and Gendered Contribution in Open Source

In open-source software development in general and on GitHub in particular, socio-demographic diversity is lower than anywhere else in the tech industry [32]. Women are particularly underrepresented, with a 2020 literature review by Trinkenreich suggesting women make up 9.8% of OSS contributors [101], a 2017 survey of open source contributors on GitHub placing them at less than 5% [39], and data mining by Canedo et al. suggesting they are even less likely to be represented among core developers on a project (2.3% core versus 5.4% overall [12]).

Women are also more likely than men to encounter stereotyping or unwelcoming language [71, 97, 105, 106]. For example, one study showed that women whose gender identities are revealed are less likely to have their contributions to open source projects accepted [97]. Another study observed biases against women in open source tools and infrastructure [66], while another identified barriers for womens' participation on an online coding Q&A site [36]. A mixed-methods empirical study of 336 GitHub contributors (half men, half women) found support for the presence of a gap between women contributors' competence in a technical sense and confidence in a social sense [110]. Another study analyzed joining patterns of new members to mailing lists of six open source projects, and found a higher dropout rate for women newcomers [57], and a survey of 119 open source contributors found that some are strongly opposed to including women [62]. Imtiaz et al. found that women are more likely to be more restrained in their online discourse in OSS projects [50] due to social expectations, in line with biases observed in the corporate world [111]. In particular, one study of implicit bias in the corporate world found that both men and women software engineers implicitly associated technical leadership with men, and further

found that this affected their decisions when evaluating technical contributions and candidates for employment [109].

Past work centers code production as the primary role of "core" developers, often using code contribution as a metric by which to determine core status in a project (when not using *committer* status, another code-centric signal, discussed above) [69, 80, 114]. However, other papers challenge this notion [40], and recent work suggests that signals of a healthy community are important to attracting new contributors [103], or that projects with good documentation are perceived as more likely to survive [104]. Crucially, ethnographic work suggests that while code is ascribed the morally charged status of a *"Good Thing"*, important non-code contributions such as documentation or community building, often labor done by and valued by women contributors, are undervalued by the men majority [71]. Qualitative research on contribution pathways of successful women in open source highlights the importance of community-centric roles that are hidden in that they do not leave a visible trace in software repositories or code contribution counts, but represent alternative ways to build a career in OSS [101].

Quantitative work shows that more gender balanced projects have higher productivity and lower turnover [16]. Our research attempts to understand how participation motivation evolves across the broader gamut of activities (i.e., not just code writing) which core contributors do to maintain their open source projects [98].

There is evidence that gender barriers affect women newcomers' project joining experiences. One study shows that women are less likely to continue contributing to projects with lower diversity of programming language experience, leading to an overall higher likelihood of disengagement for women contributors [76]. Gerosa et al. compared women and men's reported motivations for open source participation in their survey and found that social motivations like altruism, reciprocity and kinship were most prevalent for the women in the sample [102]. In another recent survey on participation in OSS, a majority of women respondents (64% of 22) described joining a project where friends or colleagues were also participating [73]. In interviews with 11 women in different open source projects, Singh & Bongiovanni found that the first person a woman interacted with had a major impact on participation and that mentors and women only spaces were important for success in the OSS community [84]. While the presence of women-only spaces may be one way to build connections necessary for social capital [26], one study surveyed 355 projects and found that only 12 projects had such spaces [85]. Another study interviewing potential OSS contributors found that women newcomers are wary about joining projects with cliques of men contributors, and that they are more excited about joining mixed gender projects [75]. Finally, another study shows that women newcomers are likely to face unique barriers, such as lower self-efficacy or lower feelings of acceptance [4]. Taken together, these studies suggest the joining process (covered in general in sections 2.2 and 2.3) and participation motivation across different stages of this process may also vary by gender.

## 3 METHODS

In order to investigate and contrast experiences of men and women in joining and becoming embedded in an open source project we conducted a qualitative interview study with 23 contributors to public projects within the PyPi ecosystem on GitHub. Our goal was to identify motivations influencing the trajectory of participation on a projectand to understand potential gender differences in the experiences of joining and participating in an open source project. In order to identify how women's motivations and experiences differed from those of men, we contrasted responses from a balanced sample of 23 interview participants of both genders (11 men and 12 women).

### 3.1 Sampling procedure

We localized our study to the PyPI ecosystem to control for known differences in community values and practices[11]. PyPI is a registry for packages written in the Python programming language and includes over 330,000 projects and 547,000 users as of 2021 [1]. It is one of the most popular open source software ecosystems with a diversity of data processing, data visualization and data science related projects as well as packages integral to data science and scientific computing workflows [74]. As a result, it attracts contributors and participation from a variety of domains (e.g., physics, geography, web development) and both academic and commercial institutions [104]. Both aspects of the community provide good variation within which to observe contributor pathways.

We began with a list of projects indexed on the PyPi repository. By linking this list to the projects' associated GitHub project pages, we then identified contributors to these projects and these contributors' GitHub accounts. We retained only accounts which had, in any given 3 month time window within the last two years of the study, had made at least 5% of the commits to one of these projects.[10] This resulted in over 30,000 GitHub profiles to choose from. We prioritized profiles who had activity within the last three months at the time of recruitment such that participants may have a more detailed and accurate recollection of recent events.

We then applied a name based gender classification tool used in past work[76]. We manually retained only candidate accounts which were not visibly bots, had contributed to projects which were not wholesale clones of other projects, and had been involved in projects with some evidence of a public community (e.g., more than one contributor, public discussion in issue or pull request threads, documentation of communication channels). We further filtered out candidate participants with no identifiable method of contact, such as a publicly listed email address, LinkedIn handle, public Twitter account, or "Contact Me" page on a publicly listed web page, as well as those which had expressly asked not to be contacted to participate in research [2]. We attempted to correct any obvious mistakes in our automatic gender classification (e.g., by comparing the automatically classified gender to the pronouns used in people's personal bios). Our initial gender classification on the 80,000 profiles we analyzed suggested that 5.24% of PyPI contributors were women, indicating this ecosystem is representative of open source projects on GitHub with their distribution of 5% women contributors as reported in their 2017 open source survey [39].

It is important to note that we relied on automatic gender classification as a way to identify individuals on the GitHub platform that were potentially likely to be viewed as women by other contributors. Given the relative scarcity of women in the open source ecosystem, utilizing automatic gender classification helped us more readily locate individuals who were likely to be women. We were mindful in our approach of the potential harm of automatic gender recognition and the likelihood for misclassification [35]. In order to accurately assess gender for each participant and follow best-practices for gender equity and inclusivity in research as noted by [82], we asked individuals to self-identify their gender if comfortable during the interviews themselves, meaning our analysis relied on participants' own gender identity rather than that assigned by the automatic gender classifier.

### 3.2 Recruitment

We contacted a subset of the resulting list of candidates to invite them to participate in an interview, focusing on those who had been active on GitHub most recently before those who had been active less recently, until we reached saturation. As noted above, we recruited core contributors on open-source software projects that are published on the PyPi ecosystem, a widely used package

[1] https://pypi.org/

[2] https://github.com/slang800/do-not-survey-list

Table 1. Interview participant demographics, background and information about the project discussed. (*) denotes participants who were the owner of the project repository they focused on in the interview, (T) denotes participants who self-identified as transgender.

| No. | Gender | Age | Country of Origin | Industry | Years of Coding Experience | Focal Project Domain | Initial Involvement Job-related? |
|---|---|---|---|---|---|---|---|
| P1 | Man | 45-54 | SVN | Self-employed | 8 | Enterprise | Yes |
| P2 | Man | 45-54 | US | Technology | 25 | Tech infra | Yes |
| P3* | Man | 25-34 | US | Self-employed | 23 | Security | Yes |
| P4 | Man | 25-34 | US | Healthcare | 20 | Package | Yes |
| P5* | Man | 25-34 | UK | University | 5 | Design | Yes |
| P6* | Man | 25-34 | BRA | University | 13 | Mapping | Yes |
| P7 | Man | 35-44 | PRT | Self-employed | 30 | Enterprise | Yes |
| P8 | Man | 25-34 | UK | University | 14 | Research | Yes |
| P9* | Man | 45-54 | US | Technology | 35 | Communications | Yes |
| P10 | Man | 25-43 | CZE | Technology | 14 | Framework | Yes |
| P11* | Man | 25-34 | ARG | Government | 7 | Mapping API | Yes |
| P12* | Woman-T | 18-24 | US | University | 10 | Security | No |
| P13 | Woman | 25-34 | US | Web Dev | 3 | Website | Yes |
| P14* | Woman | 25-34 | US | Technology | 9 | Gaming | Yes |
| P15* | Woman | 25-34 | US | Technology | 10 | Research | Yes |
| P16 | Woman | 35-44 | ARG | University | 15 | Data mgmt | No |
| P17 | Woman | 25-34 | ITA | Web Dev | 5 | Data mgmt | Yes |
| P18* | Woman | 25-34 | FR | University | 10 | PIM app | No |
| P19 | Woman | 25-34 | US | University | 10 | Geophysics | Yes |
| P20* | Woman-T | 25-34 | US | EdTech | 8 | Education | No |
| P21 | Woman | 25-34 | CAN | University | 7 | Science | Yes |
| P22* | Woman-T | 35-44 | US | Web Dev | 27 | Framework | No |
| P23* | Woman | 25-34 | CAN | University | 8 | Geospatial API | Yes |

manager for the Python programming language. [3] We defined core as contributors who contributed 5% or more of commits to a project in at least one three month time window within the past two years based on previous work.

Over several rounds of data collection, we reached out to candidates that fit our selection criteria via email to see if they would like to participate in the study and we did not offer any incentive for participation. We reached out to 40 participants that we identified as men, and received responses from 17, with 11 of those agreeing to an interview. We also reached out to 57 participants that we identified as women, receiving 23 responses with 12 agreeing to be interviewed. A chi-squared test showed there was no significant difference in response rate ($\chi^2$ (1, N=160)=0.33, $p$=0.85) or invitation acceptance across genders ($\chi^2$ (1, N=63)=0.17, $p$=0.68). This resulted in a sample of 23 interview total participants, almost evenly balanced by gender. Table 3.2 presents an overview of our participants by demographics and open source software participation experience.

12 of our 23 participants were originally from the U.S., while 5 others immigrated from their home countries to the U.S., and one participant emigrated from the U.S. to Canada. In all, participants originally hailed from 4 continents: North America (14), Europe (10), South America (3), and Asia (2).

[3] https://pypi.org/

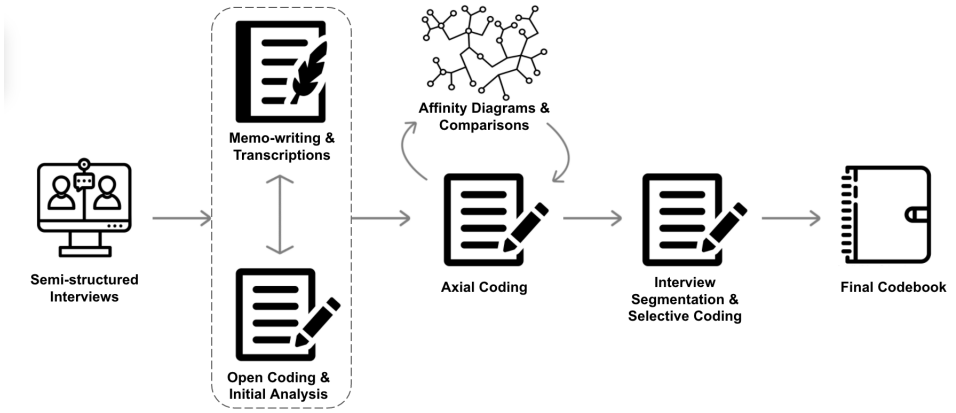


Fig. 1. Overview of qualitative coding process

16 participants self-identified as owners of the project where they had the most involvement (10 of the men & 6 of the women in our sample). 13 participants self-identified as current academic contributors (7 of the men and 6 of the women in our sample). Participants also self-identified if they contributed to the open source project we discussed in our interview as part of their employment. Interestingly, 11 of the 11 men in our sample indicating having contributed to the project they discussed as part of their employment, while only 6 of the 12 women indicating having contributed as part of their employment.

### 3.3 Interviews

We conducted semi-structured interviews to trace participants' history of participation with a specific project, as well as good and bad experiences contributing to open source more broadly. We asked participants about their background, initial involvement with open source software development, and general participation in open source software. We then asked a series of questions about their experiences with a specific focal project on which they had become a core or highly active contributor including how they initially became involved with the project, how their participation evolved over time, their current activities and involvement on the project and a set of questions about their positive and negative experiences on the project and in open source more generally. Interviews were conducted over the phone or via video-conferencing software and lasted about 1 hour on average. We transcribed audio recordings of the interviews using the automated transcription service Temi and these transcripts were subsequently corrected by the first and second authors.

### 3.4 Analysis

Our goal in our analysis was to identify initial motivations for engaging with the open source project they discussed, factors that led contributors to become more involved with the project over time and any differences between women and men on these factors. In order to identify influences on the level and nature of contributions over time, we used a Straussian grounded theory approach to data analysis, constructing theory based on the emergent themes and patterns in our data[94]. We carried out three rounds of coding activity: open coding, axial coding and then selective coding to understand and describe influences on participation over time. Figure 1 provides an overview of the research methods and interview analysis process and Figure 2 provides a working example of code development across the three rounds of coding activity.

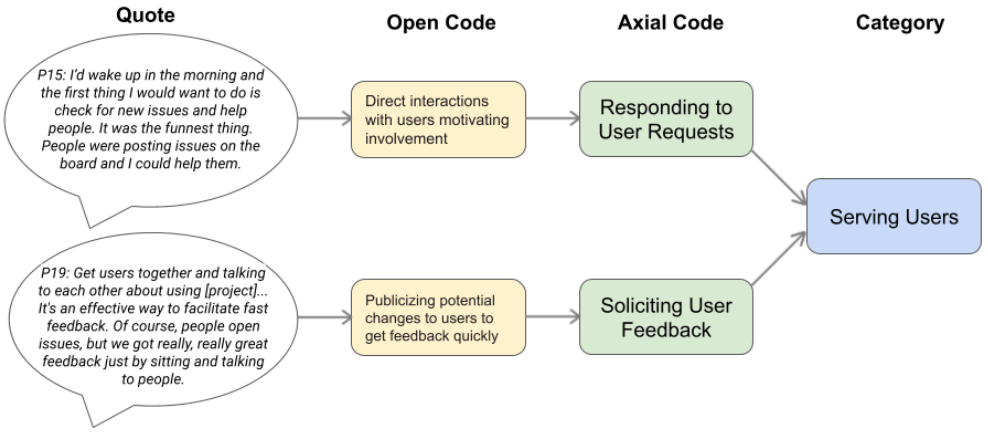


Fig. 2. Working example of coding process. Open codes, axial codes and resulting category associated with two quotes in our dataset.

*3.4.1 Open coding.* The authors began by conducting five rounds of open coding on interview transcripts and integrated memos they had written after each interview. We focused on participant descriptions of their core involvement in the project they discussed. Two authors independently open coded one interview, generating codes for any descriptions of increases and decreases in their contribution effort on a project as well as particularly positive or negative participation experiences. They then convened to discuss each coded piece of data and the corresponding generated codes through a constant comparison method, consolidating those codes into a shared set of developed codes, by combining overlapping codes or identifying new codes during discussion. Over five coding and discussion sessions, these authors generated a preliminary set of 140 open codes by independently coding the same interview, and then convening to discuss the generated codes, expanding the set of codes as additional transcripts were considered. We continued conducting subsequent interviews, memo-writing after interviews, iteratively coding and then synthesizing newly generated codes with existing codes until we reached theoretical saturation, with no new themes appearing from our dataset after 21 interviews.

*3.4.2 Axial coding.* We next performed axial coding on our full set of open codes, considering the relationship among our open codes to develop a set of initial higher level categories of developer motivations for increasing effort in the project they discussed. We used affinity diagramming to consider relationships among open codes and examine additional data from our interviews in light of these categories. Figure 2 presents examples of two axial codes generated from open codes in our analysis process and the higher level category connecting both of these codes: serving users as something that drove increased participation. The two first authors then independently coded additional transcripts using this new set of codes and then convened to discuss code application and disagreement. We applied our refined set of codes and categories to an additional set of 9 transcripts looking at instances of engagement or continued participation in a project fit into our existing

categories. We compared, contrasted, and consolidated the existing codes and sorted our memos under emerging categories, taking into consideration the differences between the experiences of our men and women contributors. Through nine rounds of code application and discussion, we further refined our codes to four categories comprising 13 codes and 66 subcodes.

*3.4.3 Selective coding.* We applied our developed set of codes to our complete dataset of 23 interviews, focusing on participant motivations and trajectories. One author segmented the interviews to identify sections of the interviews where the participant was describing: 1) their initial joining process, 2) evolution of participation, current participation and planned future participation as well as their positive and negative experiences within the project and open source more broadly. We then applied the codes describing their experiences within the projects to the data about their initial joining stories, their project participation evolution and description of good or bad experiences within a project. This code application and subsequent analysis of emergent themes revealed unique barriers and patterns of behavior within each stage of involvement as well as particular experiences associated with reduced involvement. It also supported comparison of patterns of code application across gender.

Our analysis of project participation trajectories revealed initial involvement motivations and motivations for becoming more embedded in a project, as well as the role of personal identity within each stage of involvement. Participation level was not always consistent over time, and fluctuated based on experiences in the project and personal factors. We also identified a set of detractors - factors that were associated with reduced levels of involvement on a particular project. The results of this analysis are summarized in the codebook presented in Table 2 and Table 3 in Appendix A.

In order to address our first and second research questions on motivations for initial and increased contribution to an open source project, we leveraged the gender balance in our sample to contrast the men and women participants and identify differences in their participation trajectories. We did this by first segmenting participant transcripts by stage of project involvement being discussed: initial involvement motivations, motivations for subsequently getting more involved over time, and reasons for decreasing or discontinuing involvement. We then applied the set of developed codes shown in the codebook in our appendix to each segment. Next we compared at the participant level frequency of each form of motivation based on codes applied across the men and women in our sample using Fisher's exact test to compare the count of participants with that code applied in the relevant stage [33]. Finally, to address our third research question on unique experiences faced by women in the sample, we utilized our code application by gender comparison to contrast positive and negative experiences in open source in order to identify experiences unique to the women in our sample.

## 4 RESULTS

Our analysis revealed important differences between the women and men in our sample in terms of their initial motivations for engaging with the open source project they discussed and the factors that led them to become more involved with the project over time. Figure 3 illustrates the initial and continued involvement motivations observed in our sample. Bold lines in the figure indicate motivations more prevalent among the women participants we interviewed. Our results suggest that women were more likely to engage with a project for personal reasons (personal interest, learning goals or invitation by a friend or mentor) rather than job related ones. They were drawn to contribute more as a function of personal relationships built with other developers as well as motivation to serve the community of users, that is, other developers whose code relied on the project. We describe each of these results in more detail below.

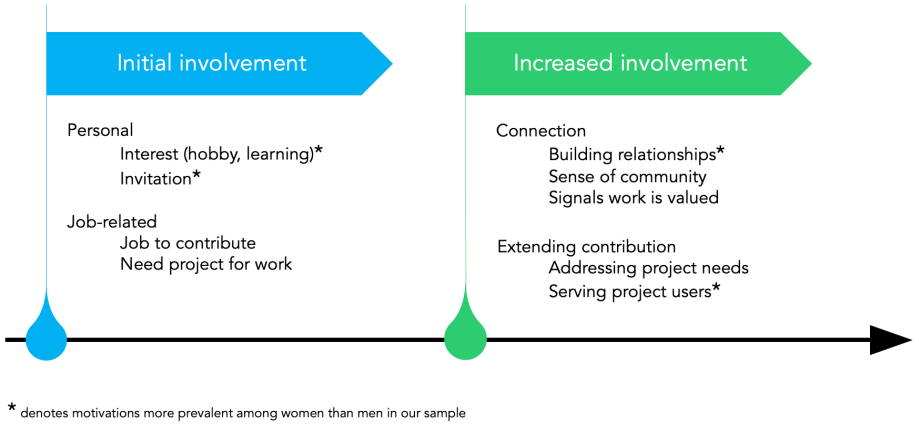


Fig. 3. Motivations for involvement with a project over time. Asterisks indicate motivations more prevalent among the women contributors in our sample as compared with the men.

### 4.1 Initial Involvement Motivations

Our first research question (RQ1) was: **What motivates women developers, as compared with men, to initially engage with an open source project?** We addressed this question by contrasting the men and women in our sample in terms of their initial motivations for engagement with the open source project they became embedded into. We observed a core set of initial involvement motivations consistent with those described in previous research on open source: creating or extending a project to address job needs, being paid to contribute, creating or using a project for learning and self-development, and creating or extending a project to address a personal need [38, 102]. Table 2 in Appendix A summarizes our full set of codes around initial involvement motivations.

Although other research on joining open source projects has described the fact that people join projects as part of their work or out of personal interest, we found these motivations were not equivalent across the men and women in our sample. The left side of Figure 3 overviews initial joining motivations, highlighting in bold those that were more prevalent for women in our sample.

*4.1.1 Job-Related Motivations for Initial Contribution to Project.* A major reason why both men and women participants in our sample started contributing to an open source project was as part of their paid employment. In such cases, the participant's job involved writing code, which involved either: 1) contributing to a project already used by or created by their institution or organization, 2) seeking out a project that filled a technical gap needed as part of the participant's job, or 3) creating a piece of software needed for their work and deciding to open source it.

One of the most common job-related motivations for initially contributing to a project was debugging, or finding and fixing flaws in the code base, as a project user. As participants started off using the project they realized there was either a feature or capability that needed to be added or a bug that needed to be fixed. They submitted a contribution to the project in the form of a pull request, or PR, and this led to further involvement once it was accepted. For example, P4, a

man in our sample, wanted to use a project for his own use and identified some missing features and unsupported functions. Noticing an obvious gap, P4 was spurred to submit PRs to remedy the situation:

> *So <Project2> had a lot of that in place, so I thought it would be easier to just use that. But it was still missing features here and there in terms of which way the memory layout is in the arrays. (P4)*

Some developers were directly assigned to contribute to an open source project as part of their job, acting as a maintainer for an organization owned project or contributing to a project their organization relied on for their own code. Some commercial institutions sponsor open source projects for various reasons, and many companies have projects completely on Github. P13, for example, initially got involved in the open source project she worked on through a work assignment. She had already been working within her company for some months and was assigned to the project (building a website for a company client based on an existing website from a Django project) because she was the most available at the time.

For the specific project we focused on in our interviews, the men in our sample were significantly more likely to get involved with a project for job-related reasons, creating or extending a project to address technical needs in their work or employed specifically to maintain or contribute to the project (11/11 men vs. 7/12 women in our sample - Fisher's exact test statistic = 0.0373, $p < 0.05$).

It is important to note that for some of the men and women in our sample where contributing to open source was motivated by job needs, they didn't always see it as part of their formal job description. This was particularly true for academic contributors. As research on scientific software development has pointed out, open source contributions, and software development more broadly, isn't considered part of the scientific process or academic research activity(e.g. [48]). For example, P23, a woman researcher in our sample described getting pushback and discouragement from contributing to open source from the faculty within her department. They did not consider her open source development to be research and told her she was "wasting her time."

*4.1.2 Invitation.* Another reason participants got involved initially with a project was through their social network. Several participants were invited into a project by a mentor, a colleague, an acquaintance, or a friend. Through word of mouth and these connections, the participant became aware of the existence of the project if they were not already, and were exposed to its merits and applications. Their friend or colleague on the project encouraged them to check it out or directly asked them to contribute, leading to their eventual involvement. For example, P21 was a university postdoctoral researcher in physics. Her postdoctoral research advisor was a major contributor in an open source data science project and invited her to contribute. As she described:

> *And whenever I graduated it was just like a connection thing. You know, one of my mentors put me in touch with my current boss and they were kind of impressed with the work that I had done on <Project1> and since <Project2> is kind of similar, they wanted me to just kind of continue that kind of work. (P21)*

A larger proportion of women in our sample than men joined the project they discussed as a result of a direct invitation from a friend, colleague, or mentor (5/12 women vs. 1/11 men in our sample - Fisher's exact test statistic = 0.155). While invitations did not appear to be a motivating influence for our sample of men, multiple women expressed being invited was part of their rationale for joining a project. Although this difference was not statistically significant, the influence of a trusted recommendation coming from a familiar individual is mentioned in other research on women and open source software participation [100].

*4.1.3 Personal interest.* For several of the developers in our sample, initial entry into the open source project was driven by the desire to pursue a new interest or to develop or strengthen new skills, outside of their assigned work or studies. The participant's interest in the project, in these cases, was usually one of a technical nature to satisfy a personal passion or curiosity, to broaden their technical skills, or to increase their employability through contribution. For example, P20 was already contributing to other open source projects but cited a desire to develop their skills and to learn a new programming language, Rust, as their motivation for contributing to a new project:

> *I'm interested in learning Rust. Um, so I've been helping out with a shell in Rust Called <Project> (P20)*

Women cited personal goals of actively trying to learn a new skill and wanting to build something on their own more often than men (5 out of 12 women and none of the men in our sample - Fisher's exact test statistic = 0.0373, $p < 0.05$). Women wanted to build up their technical skill set and enhance their personal projects and reputation. This may be in part because the women in our sample were often working in non-technical fields or fields outside of software development (e.g., Geophysics).

## 4.2 Motivations to Continue Involvement

Our second research question (RQ2) was: **What motivates women developers to increase their involvement over time?** We identified a key set of motivations for participants to become more embedded in their project and deepen involvement. These motivations fell into two broad categories: 1) taking on more responsibility in the project through different forms of work, and 2) becoming connected with the project through contributions and interactions. Table 3 in Appendix A overviews these categories of motivations and associated codes identified in our analysis. A subset of these motivations were more prevalent for women than men, as described below.

*4.2.1 Different forms of work.* As participants became more involved in the organizational structure of the project, their behaviors and responsibilities changed. They engaged in a wider variety of project activities beyond simply writing code, including taking on management responsibilities, responding to issues, reviewing code contributions, participating in design discussions and decision-making about project direction over the longer term, and engaging in community building activities. In becoming more involved, our respondents reported actively identifying gaps in project structure or other project needs, and with their additional responsibility and reputation, taking a role in making decisions to address these gaps. Participant engagement, then, related to perceived gaps in the existing project and community structure was aimed at filling these gaps by increasing their involvement with a specific form of work.

We found the men and women in our sample varied in the nature of the responsibilities they took on once they were an established member of a project. The women in our sample were significantly more likely to engage in and be motivated by serving a project's user community (10 out of 12 women vs. 1 out of 11 men in our sample - Fisher exact test statistic = 0.0272, $p < .05$). Both men and women engaged in addressing project needs through code and contribution management and development of social infrastructure (community spaces for communicating among members), and project management activities such as managing design decision discussions.

Women in our sample uniquely discussed caring about and being energized by supporting the community of developers surrounding a project through responding to their requests for help in issues, taking their concerns into account and advocating on their behalf in project level decisions. They described how it was exciting to be able to respond to users' needs and interact with them

directly to learn about their concerns. These interactions often began with a user query that spurred the participant into helping users solve a problem they've reported, helping them get what they need, and responding to user questions, issues, or messages on a variety of forums such as Slack, email, or Stack Overflow. P15 described how it was motivating to respond to issues and help users:

> *I would wake up in the morning and the first thing I would want to do is check for new issues and help people. It was the funnest thing. People were posting issues on the board and I could help them. (P15)*

The women developers were also strongly motivated by general concern about their project's users and their needs. Here, P19 describes how the rest of their team was concerned with performance and they had to advocate on behalf of the users' that usability was a more important area to be worked on.

> *I'm often arguing that [performance] is not important to our users and what is important to them is usability and how easy it is to actually run things. And so with limited manpower, we need to focus on usability instead of performance. (P19)*

This quote exemplified several from our women participants talking about the importance of project usability. They frequently described making design or technical choices with existing users in mind, and making a conscious effort to consider how decisions or actions would affect their users. While the men in our sample also discussed interactions with users and considerations on usability, these were due to direct issues that appeared. The men in our sample did not exhibit the same forward-looking consideration of community or user empathy (i.e., putting oneself in the mindset of the user) as the women did.

*4.2.2 Becoming core through connection: Valued work and social connections.* The social behavior surrounding a participant's contribution to the project code base had a major influence on continued participation. Contributors became more involved through building relationships with other developers on the project. A sense of community and belonging was also motivating and sustained participation. Positive interactions around code contributions made participants feel their work and efforts on a project were valued. These interactions could include their code being used by other developers, getting feedback on their contributions, and generally feeling as though their work was valued by the community. Table 3 in Appendix A summarizes these social influences on increased participation in a project.

In contrasting these social connections across men and women, we found that women were more likely to have increased contribution as a function of building relationships with other developers on a project (9 out of 12 women in our sample and 3 out of 11 men - Fisher's exact test statistic=0.0391, $p < .05$). Extended social connection to the project involved interacting with other project contributors across more and more communication channels outside of the GitHub site itself (e.g., Slack, text messaging, and phone) as well as meeting in person. P15 described how they initially became a member of the project by joining the mailing list. They then started interacting with other developers, eventually speaking to the lead developer about filling in a role that P15 already had experience in.

> *I think probably the first thing that I know I joined the mailing list and I know that then I started talking, I actually had a phone call with the lead developer and I was like, look, we need a registry person for <Project> containers and Hey, I'll build it. (P15)*

This was common across many of the women in our sample, who described an increasing number of interactions with developers on their project through the phone, text messaging or the project Slack as they became more and more involved. They began to form friendships with the other developers on the project through these interactions.

In-person meetings with other contributors were critical events where developers formed friendships with other members and collaborated. At these events, participants had interactions with other project developers, submitted contributions, and became more tied to the project. While both men and women in our sample described moving on from their initial uncertainty when becoming more involved on a project, women were more conscious of their presentation and role to others. P23 described being very shy and 'very quiet' at her first remote meetings due to her novelty and self-perceived incompetence.

> *I was actually going to the monthly <Project1> development meeting and at the first I was very quiet because I didn't know what they were talking about but as time [went by] I could speak more.(P23)*

Although P23 started out very unsure of her role and expertise initially, over time, she described becoming more comfortable with the subject and speaking up more and more often at the meetings. She eventually met the other developers face to face and found that they were 'very helpful' and 'super nice.'

### 4.3 Womens' Unique Experiences in Each Stage of Participation

Our third research question (RQ3) was: **What experiences in open source are unique to women developers?** In contrast to most of the previous work on open source, our study included a balanced sample with a fair number of women participants (versus focusing only on men or only on women). This gives us an opportunity to compare our participants' experiences, and in doing so we note that women faced a unique set of pressures because of their rarity and associated visibility within the open source environment. Women in our sample described additional pressure they felt to prove their competence as the only woman on a project, facing excessive scrutiny on their contributions, and gender non-inclusive language and behavior that signaled they were not supposed to be in the environment.

*4.3.1 Early Stage Participation Challenges.* When initially joining a project and in the early stages of participation, women in our sample described additional pressure they felt to prove their competence as the only woman on a project. The transparency and public visibility of open source software development activities creates audience pressures for developers as their code contributions and discussions around code are readily available for thousands of others to see and comment on [22]. The women in our sample were acutely aware of this pressure and the fact that the audience for their work was predominantly men. Issues regarding visibility and self-presentation impacted women's behavior upon entry into the project, and could persist as their participation continued.

*Visibility Pressures:* Four women remarked on the fact that they were the only woman on a large project. They described how being the only woman in a particular project to them meant a higher level of pressure to prove themselves when first engaging with a project. There was an awareness that as a woman contributor, the bar and expectation for contribution may be higher, associated with increased levels of scrutiny on their contributions. P23 described this experience in the project she worked on and how she limited her contributions because of this pressure:

> *For a long time I was the only female [on my projects]. So most of the time I was uneasy and felt under pressure to prove myself. So in the first several years I was just quiet and I tried to learn as much as I could in my spare time.(P23)*

Fear of being attacked solely for their identity was a potential deterrent for woman, trans, and non-binary involvement and motivated actions to obscure or minimize visible indicators of gender. This is akin to findings on fear of gender-based attacks in Wikipedia[67]. P15 in our sample hid her

identity on the Slack channel for her project so that her gender was not immediately identifiable to other contributors and users by using a single letter alias rather than her full name.

> *I feel like if I had something like [NAME], it would identify me right off the bat as a girl and I'll get more respect if I don't do that.(P15)*

P20 discussed how she and her co-maintainer, both trans open source contributors, are aware of and preemptively expect the attacks they may receive as their project becomes more popular. The increased popularity and visibility of the project brings the identities of its developers to light as well, which the maintainers were aware may run the risk of exposing them to unwanted negative attention.

> *I know my co maintainer and I are both trans folk so we are very aware that as we get more popular, like that heat is coming… successful women and successful trans folk on the internet often get abused in various ways. We are both aware that this thing is probably coming one day. (P20)*

These quotations speak to a preemptive defense strategy from women contributors who were aware that women may be attacked online on the basis of their gender, and as a result chose to keep their identities hidden for as long as possible and remain vigilant of any opportunities for attack.

*Fear of Looking "Stupid":* Three female respondents expressed a fear during their initial interactions with a project of asking questions or being highly visible and not knowing all the details about a project without full confidence they knew what they were doing. This was particularly a concern early on in engagement with an open source project as contributors were first orienting to the code base itself and made assumptions that the questions they had were obvious to more experienced contributors or trivial in nature. They expressed a fear of public missteps in both comments and actions on the code itself that would expose a lack of knowledge or make others think they did not know what they were doing:

> *Sometimes I'm a bit afraid to make a pull request on some other people's project just because I don't know if I'm doing this right...I guess it's a fear to be wrong. (P18)*

When discussing her early days in open source, P17 described help-seeking as a negative behavior. She took pride in becoming expert enough that others would ask her for help and that she could now be "useful" to the project and work without needing to ask for help. She described her early days of continually asking questions as burdensome on the main developer:

> *You work and you don't know if what you're doing is good enough. You feel like you cannot ask questions, or ask as many questions because if you're asking that many questions it means you're not capable of solving problems. (P17)*

She described experiencing imposter syndrome early on due to coming from a different background and overcoming it by having her work used by others.

> *Personally, switching from bio engineering to software engineering, there's a lot of imposter syndrome that comes with that. And so knowing that now a piece of my work is out there and people are actually looking to me for answers on how to use it. It just felt nice.(P17)*

Positive contribution experiences such as this one and external recognition as an expert helped the women in our sample to overcome this pressure and fear of looking stupid. As the women developers in our sample became more established they were less concerned about this kind of pressure and more confident in their abilities as an expert based on their contribution history.

*4.3.2 Continued Engagement Participation Challenges.* As the women in our sample continued participating in a project, they encountered a different set of negative experiences related to regular

interaction and sustained contribution than those of men. They described experiencing excessive scrutiny on their contributions to the project, both in terms of code submissions and in discussions about project decisions or direction, encountering gender non-inclusivity in the language and actions in project discussions, and having to discontinue participation because of maternity leave or external pressures.

*Excessive Scrutiny:* In continued engagement with a project, women respondents described experiencing excessive scrutiny of their contributions and ideas. This was akin to the 'superfluous labor' described by Singh and Bongiovanni [84] where their participants described having to put in extra effort to make a case for the accuracy, legitimacy and value of a piece of work. Participants perceived a difference in the level of effort required to justify their work or ideas in contrast with the men on the same project or in contrast to how they were treated in other communities. P15, a regular contributor and core member on her project, described noticing this excessive scrutiny in terms of relative numbers of comments on her contributions during pull request review within a particular project. Because she participated on multiple projects she was aware of a clear difference between her treatment within this project and the others she worked on. As she described:

> *I started seeing this pattern where if anyone else in the organization opened up that pull request, it would get a LGTM [Looks good to me] Maybe one or two comments and then it'd go in. With my pull requests for a particular community there would be, I literally counted like fifty, up to a hundred, comments and nitpicking on things that didn't matter. And then people telling me how to rewrite the code.(P15)*

This experience conveyed a lack of trust in her competence and ability within the community. P20 described a similar experience when presenting her work at an open source conference to another developer, who challenged her decisions aggressively. Because of her level of establishment she was not intimidated by this questioning but recognized it could be difficult for a person with less experience. As she described:

> *And somebody started digging in. Like, why would you make this decision? Because these are not the right ways to do engineering.(P20)*

*Gender Non-inclusivity:* Our women participants also encountered gender non-inclusive language or behavior. For example, one participant described developers on her team using sexist emojis in the project Slack channel and equated it to the "bro" culture behavior she had witnessed at technology conferences. Another participant noted some gender non-inclusive language within the workplace. One of her coworkers on the project referred to experienced and knowledgeable colleagues as "greybeards," implying that only old men may ever get to such a stage. As P19 described:

> *There was another developer who whenever he's talking about people who have been around and really know what they're doing, he often refers to them as gray beards, which I don't find that particularly inclusive. 'Cause it's like the only people that could've been around a long time and know what they're doing are men.(P19)*

Non-inclusive language and behavior reinforced the idea that developers were predominantly men and only needed to cater to the other men in the environment with their behavior.

*Maternity Leave:* Three of the women in our sample went on maternity leave and had to reduce or discontinue completely their involvement in the open source project they discussed. Maternity leave has a significant impact on a woman's availability and project involvement, and made continued engagement with open source more difficult. One of our participants was even breast-pumping during our virtual interview. It was difficult to stay completely disconnected during this time because of the ease of digital access. Our participants spoke about leaving their project temporarily

during their maternity leave (12 weeks), but staying involved and updated with their respective projects through Slack, email, etc. There was also an awareness that the extended period away would hurt their visible contribution history in open source as P19 described:

> *I did take a maternity break, a kind of extended one where I was more or less gone for six months. So there's kind of a big gap in contributions there. (P19)*

While participants obviously contributed less to their projects while on maternity leave, they were still connected enough to the project to actively continue working and remain updated, despite not being required to work on it.

## 5 DISCUSSION

In this study we describe differences in motivations for women, as compared with men, across stages of joining an open source project and unique contribution experiences for women contributors in open source software development projects in the PyPI ecosystem. While both men and women in our sample got involved or initiated a project because of job related technical needs or job assignments, women were more likely to be drawn in by personal interest or direct invitation from an existing heavily engaged contributor (RQ1).

As participation increased, women were more motivated by serving project users' needs through responding to their requests and making the project more usable (RQ2). Both men and women increased participation after getting signals that their contributions were valued by other developers and project users and because of a sense of community in the project, but women were more likely to become embedded through building relationships with other developers on the project. As their participation continued, both men and women got involved in a greater variety of different forms of work, increasing participation to address project needs through project management activities, code and contribution review, and building and maintaining social infrastructure such as community building, mentorship activities, and project advocacy.

Women faced a set of challenging experiences as contributors because of their visibility as unique members in the space (RQ3). They described feeling pressure as a result, excessive scrutiny on their contributions, fear of looking stupid or possible gender-based attacks, and a gap in contribution during maternity leave. Here we consider implications of our results for the study of open source software development participation more broadly as well as the survival and health of open source projects. Finally, we consider the implications for diversifying participation in open source.

### 5.1 Invisible and Alternative Forms of Work

Our results highlight the wide set of project maintenance behaviors beyond writing code that drew contributors in and sustained their involvement as embedded members of a project. Women in our sample were more motivated to increase engagement with a project to support the community of users around a project. Our results connect with recent work on the hidden labor of maintenance activities [77] and the need for project advocacy behaviors often engaged in by women contributors [100]. Women may be more likely to act as 'hidden figures' supporting a project through community engagement [101]. There is increasing evidence that devoting time to project advocacy, serving users, and social infrastructure are other valuable forms of work representative of a project core member.

Non-coding forms of work are often difficult to extract when mining GitHub repositories for data, e.g. the Slack channels or offline venues where community building and user support addressing project needs is carried out. Future work should consider ways to measure and track some of the hidden labor associated with project advocacy, serving users, and creating and maintaining social infrastructure.

It is not clear whether the "invisible" work of project advocacy and community support is evenly distributed across types of contributors (i.e., women and underrepresented minorities). Research in other settings suggests women take on a disproportionate share of so-called non-promotable work [2] and in the home carry out a great deal of invisible labor to support their families [6]. This may also be the case in open source. Given that a majority of previous empirical studies of open source participation has focused on code contribution metrics to define core members, it is possible that many have missed out on or mischaracterized women's participation. Future work should investigate and analyze how not only work is distributed across different types of participants in a project, but also the different kinds of work (i.e., technical v. non-technical contributions). Open source projects and open collaboration platforms such as GitHub need to consider ways to capture and include metrics of these different kinds of work to accurately represent contributors' participation and recognize non-technical efforts.

### 5.2 Gender Inclusiveness

Our results suggest women face unique obstacles in open source communities (e.g., increased visibility pressures, less confidence, feelings of scrutiny), making it more difficult to initially attract them as contributors and convince them to remain. For these reasons, we encourage early intervention to help build confidence in newcomers, especially from under-represented or vulnerable groups. By highlighting the unique challenges faced by women in this environment, we encourage open source communities to be aware of the issues women face and their experiences so they can actively work towards making their projects more welcoming to all. Doing so may not only encourage initial participation from newcomers, but provide the foundation for retention and community-building to support long-term contributions from women contributors and contributors from other under-represented groups.

Despite numerous studies, reports, and recommendations for broadening computing, women are persistently underrepresented in computing careers.[21][65][17][83] According to the National Center for Women in Technology (NCWIT), from 2003-2017, while women comprised 56 - 58% of the overall workforce in the US, they made up only around 20% of the computing workforce.[24] Only 17.9% of bachelor's degrees in computer science were awarded to women in 2016 [117]. The research on barriers to women's representation in technology careers suggest three important potential barriers: self-efficacy, culture, and contributor social capital [17][83]. Our results have implications for addressing each of these types of barriers.

*5.2.1 Self-Efficacy.* The women in our sample expressed lower levels of confidence in their abilities, or self-efficacy, when first joining a project and fear of looking stupid by asking questions or seeking help early on. Self-efficacy is the belief that one has the capacity to be successful at a task, as it influences motivation and effort[5, 7]. Confidence in computing abilities is a major influence on selecting and staying in a computing major, shown across a range of quantitative and qualitative studies on attrition in undergraduate computer science[8, 18, 54]. Inaccurate self assessment, stereotype threats, and limited exposure to computing experiences all contribute to this lower level of self-efficacy.

Women are particularly susceptible to underestimating themselves and therefore having lower levels of self-efficacy and self-confidence[112]. It is therefore important to encourage and support women newcomers in any way possible. The most direct way is to welcome and guide the participant through their contributions to the project. By recognizing and welcoming newcomers, community members can build the foundations for important social relationships and foster new project growth. These connections can help enhance newcomer self-efficacy and self-confidence by providing guidance and resources. As we've seen from our participants, positive experiences and familiarity

with the code base over time contribute to building self-confidence, which in turn encourages independence, creativity, and efficiency.

For open source projects, our results suggest the potential for support through private mentorship akin to that explored by Ford et al.[37] to combat initial contribution barriers for women by providing a safe space for asking questions and a guardian to walk them through the contribution and community process. A challenge women face in the computing profession more broadly is lack of mentorship, leading to early dropout. Mentoring involves a specific kind of relationship between a person with more experience (a mentor) and a person with less experience (a protégé).[56][49] Mentoring contributes to learning, professional development and career growth.[60][27] With early intervention through private mentorship, women may feel more comfortable as they have a casual and easily accessible source of help within the community. They could speak more candidly in a private conversation with a mentor and ask for advice and questions relevant to their personal struggles, as opposed to publicly asking for aid in Slack or email chain where women are subject to social expectations.[50] A sense of trust, stability, and comfort are integral for making any newcomer feel welcome. Private mentorship would require an additional burden on more senior members of the project, but the investment would be worth it as it would encourage all, not just women, newcomers to feel at ease in a new environment and subsequently encourage them to stay and contribute.

*5.2.2 Culture and stereotypes.* Cheryan et al. discuss culture as a critical determinant of why some STEM fields are more gender balanced than others, noting that masculine cultures signal a lower sense of belonging to women versus men.[17] They define culture as "a dynamic system consisting of individual behaviors and psychological tendencies that influence and are influenced by historically derived ideas and values, everyday interactions and societal structures" (p. 2). Previous work on the representation problem suggests the masculine culture of computing has a direct negative influence on women's identification with the profession and indirect negative influence through stereotyping, and biased and discriminatory behavior toward women in the field [71]. The majority of open source communities are man-dominated with occasional aggressive behavior and sexualized language towards women, which may be discomforting to new women contributors.[26]

A potential way to counteract computing stereotypes and create a more open, welcoming, and gender-equitable community could be to reinforce the diversity of identities of members of the computing profession. Developer profiles that highlight underrepresented individuals could reshape the image of developer identity. Liu et al. studied the #ILookLikeAnEngineer hashtag movement observing how women engineers sharing images of themselves on Twitter increased their feeling of belonging in the profession and empowerment [64]. Visible profiles describing the history of participation and involvement of prominent women or Black developers promoted on GitHub itself could reshape perceptions about membership in the community and help underrepresented individuals feel a greater sense of belonging. Broadening the perception of what it means to be a software developer in open source could reduce some of the pressure on women contributors as the stereotype of the computing profession as a whole becomes more inclusive and diverse. At the same time, our results suggest this kind of intervention must be done with great care. Within our sample, there were people who hid aspects of their identity because of fears of harassment. In addition, there may be some cold start problems where if only a small number of women, for example, make themselves visible they will be targeted. GitHub themselves have avoided this issue by posting profiles of very established women. Future work could examine the best way to make this recommendation actionable in a safe way.

*5.2.3 Social Capital: Building Connections Among Contributors.* Our results suggest that social capital is incredibly important for attracting and retaining women contributors. Connection with

individuals already part of an open source project may facilitate initial involvement for women contributors that are less likely to engage through their work. We found that invitation was a very common method for women to get initially involved in a project and facilitated private mentorship interactions. Women may feel more comfortable confiding in, asking questions to, getting advice from an existing connection within a project. Similar to having a mentor, knowing someone on a project helped women contributors overcome initial hurdles and helped them feel more relaxed in the project setting. Women participants who enter projects without connections may conversely feel like an outsider upon initial entry.

Building relationships with other developers also motivated women in our sample to increase involvement with a project (as compared to the men). This result echoes findings by Qui et al showing that social capital has a stronger influence on the continued participation of women developers in open source [75]. This also ties in with the existing literature on women's barriers and motivations [100]; community, kinship, and mutual connections within the user community were all important aspects of women's continued participation in a project. These are important motivators that specifically combat feelings of lack of peer parity, toxic culture, impostor syndrome, and community reception issues.

Regardless of their initial contribution motivations, connections with other community members sustained womens' project participation, even if for example their initial contribution was job-related and the project became less relevant to their work. These findings suggest projects should consider ways to foster connections between developers on a project through interventions such as those used in traditional organizations to build social networks (e.g., peer mentoring or paired project assignments). Izquierdo et al. recommend for community-wide recognition of the importance of diverse and inclusive teams, continued support of working groups, and continued tracking of female participation to facilitate parallel mentorship. [51] In addition, this finding reinforces the importance of analyzing developer social networks in thinking about contributor diversity and contributor retention.

### 5.3 Limitations

As with any research, our methods have associated limitations which may reduce the potential generalizability of our findings. In our study, we took a qualitative research approach, focusing on participants' verbal accounts of their activities and experiences on the projects they engaged with versus tracing their history through activity logs. We did this because we were interested in understanding their motivations and perspectives on their contribution history and identifying aspects of project participation that may have occurred outside of GitHub. The tradeoff is that we are reliant on self-reported data which is subject to bias or imperfect recall. In addition, we collected data from a relatively small sample. Future work that takes a quantitative survey or log data analytic approach can examine the extent to which our observations are validated in a larger sample and by activity data.

Our sample was not globally representative – the majority of our participants hailed from Northern America (14 out of 23) and Western and Northern Europe (5 out of 23), and experiences, background and cultural attitudes may vary in these other regions. Our email reaching out to participants asking for an interview was in English and, as a consequence, excluded participants who did not have working fluency in English, reducing the potential global reach and excluding many people in the community. The countries of origin for our participants, however, are typical of many similar studies on open source[95][105]. Future work should examine a more globally diverse sample of contributors to understand whether these experiences are consistent across national backgrounds and explore how national culture might interact with open source project cultures.

Another potential limitation of our sample is that it was drawn from one specific community within open source, PiPy, which contains many data processing, data visualization and data science related projects and as a result attracts contributors and participation from certain domains (e.g., physics, geography, web development) which do not wholly reflect all project domains within open source. Many other ecosystems exist (e.g., JavaScript, Linux), but only PiPy, within Python, was the subject of our study. Future work should examine diversity at the ecosystem level and expand to include other ecosystems such as Javascript which may include more front-end web developers with different backgrounds, expertise, and professional community dynamics that could influence participation experiences.

Additionally, we focused our study primarily on gender diversity and we did not explicitly seek out participants from other minority groups. There is a need to consider intersectionality in this space, and how gender, ethnicity, class, and race could influence participation experiences. There is almost no research we could identify on the issue of racial identity in open source, representing an important and needed area for future research. Our sample did include three trans-women, and our results on identity connect with some themes in previous work on inclusive development work environments for trans developers [23]. Future research could examine the experience of trans, non-binary, genderqueer, and gender non-conforming developers in open source specifically with a larger sample as a complement to the previous work in a more traditional software development work environment. Understanding intersectionality and the impact of various personal characteristics on open source contribution is essential to providing tailored intervention strategies for encouraging long-term participation.

In recruiting our sample we identified developers as core based on the definition from [10] which focuses on code contributions only as noted above. We attempted to account for this by confirming their participation levels in the project we initially used to identify them, and focusing our interviews on the project where they had most actively participated. It is possible that other measures of embeddedness or centrality within a project would yield somewhat different accounts of joining patterns and participation experiences. In follow-up work, we plan to examine participants' social network position in the broader project ecosystem to understand how individuals become embedded in the ecosystem as a whole.

In future studies, we will be examining the unique obstacles men and women face with a focus on how it affects their confidence and self-perceived competence. As shown in this study, a participant's self-confidence, relationship to others, and involvement in the code base are critical factors in maintaining involvement. We will be exploring these themes in further detail with a similar empirical qualitative study.

## 6 CONCLUSION

In this study, we analyzed the similarities and differences in what motivates women versus men to initially get involved in an open source project and increase involvement over time. While men and women entered projects for many similar reasons, women were more likely to engage for personal reasons and through invitation. Men and women in our sample took on many of the same forms of work, but women were more motivated to increase and sustain contribution to address the needs of the user community surrounding a project. Finally, women faced a unique set of challenges in their contribution experience as a function of their visibility and relative rarity in the open source community. Open source maintainers and corporations with a vested interest in the sustainability of open source should consider these results when thinking about ways to increase diversity and inclusion in their open source projects.

## ACKNOWLEDGMENTS

Intentionally left blank for anonymous review.

## REFERENCES

[1] Guilherme Avelino, Eleni Constantinou, Marco Tulio Valente, and Alexander Serebrenik. 2019. An empirical investigation of the abandonment and survival of open source projects. In *Empirical Software Engineering and Measurement*.

[2] Recalde M. P. Vesterlund L. Weingart L. Babcock, L. 2017. Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review* 107, 3 (2017), 714–747.

[3] Richard Bagozzi and Utpal Dholakia. 2006. Antecedents and purchase consequences of customer participation in small group brand communities. *International Journal of Research in Marketing* 23 (2006), 45–61. https://doi.org/10.1016/j.ijresmar.2006.01.005

[4] Sogol Balali, Igor Steinmacher, Umayal Annamalai, Anita Sarma, and Marco Aurelio Gerosa. 2018. Newcomers' barriers... is that all? an analysis of mentors' and newcomers' barriers in OSS projects. *Proceedings of the Conference on Computer Supported Cooperative Work* 27, 3-6 (2018), 679–714.

[5] Albert Bandura. 1997. The nature and structure of self-efficacy. *Self-efficacy: the exercise of control. New York, NY: WH Freeman and Company* (1997), 37–78.

[6] Albert Bandura. 2019. Invisible Household Labor and Ramifications for Adjustment: Mothers as Captains of Households. *Sex Roles* 81 (2019), 467–486.

[7] Albert Bandura and Daniel Cervone. 1983. Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology* 45, 5 (1983), 1017.

[8] Sylvia Beyer, Kristina Rynes, Julie Perrault, Kelly Hay, and Susan Haller. 2003. Gender differences in computer science students. In *Proceedings of the 34th SIGCSE technical symposium on Computer science education*. 49–53.

[9] Christian Bird, Alex Gourley, Prem Devanbu, Anand Swaminathan, and Greta Hsu. 2007. Open borders? immigration in open source projects. In *Proceedings of the International Workshop on Mining Software Repositories*. IEEE, 6–6.

[10] Nagappan N. Murphy B. Gall H. Devanbu P. Bird, C. 2011. Don't touch my code! Examining the effects of ownership on software quality. In *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering*. Association for Computing Machinery, New York, NY, USA, 4–14. https://doi.org/10.1145/2025113.2025119

[11] Kästner Christian Herbsleb James D. Bogart, Chris and F. Thung. [n.d.]. When and how to make breaking changes: Policies and practices in 18 open source software ecosystems. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 30, 4 ([n. d.]), 1–56.

[12] Edna Dias Canedo, Rodrigo Bonifácio, Márcio Vinicius Okimoto, Alexander Serebrenik, Gustavo Pinto, and Eduardo Monteiro. 2020. Work practices and perceptions from women core developers in OSS communities. In *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. 1–11.

[13] Gerardo Canfora, Massimiliano Di Penta, Stefano Giannantonio, Rocco Oliveto, and Sebastiano Panichella. 2013. YODA: Young and newcomer developer assistant. In *Proceedings of the International Conference on Software Engineering*. IEEE, 1331–1334.

[14] Gerardo Canfora, Massimiliano Di Penta, Rocco Oliveto, and Sebastiano Panichella. 2012. Who is going to mentor newcomers in open source projects?. In *Proceedings of the Joint Meeting on Foundations of Software Engineering*. 1–11.

[15] Marcelo Cataldo, James D Herbsleb, and Kathleen M Carley. 2008. Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity. In *Proceedings of the International Symposium on Empirical Software Engineering and Measurement*. 2–11.

[16] Gemma Catolino, Fabio Palomba, Damian A Tamburri, Alexander Serebrenik, and Filomena Ferrucci. 2019. Gender diversity and women in software teams: How do they affect community smells?. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 11–20.

[17] Sapna Cheryan, Sianna A Ziegler, Amanda K Montoya, and Lily Jiang. 2017. Why are some STEM fields more gender balanced than others? *Psychological Bulletin* 143, 1 (2017), 1.

[18] J. McGrath Cohoon and William Aspray (Eds.). 2006. *Women and Information Technology: Research on Underrepresentation* (1 ed.). Vol. 1. The MIT Press. https://EconPapers.repec.org/RePEc:mtp:titles:0262033453

[19] Kevin Crowston and James Howison. 2005. The social structure of free and open source software development. *First Monday* 10, 2 (2005).

[20] Davor Cubranic, Gail C Murphy, Janice Singer, and Kellogg S Booth. 2005. Hipikat: A project memory for software development. *Transactions on Software Engineering* 31, 6 (2005), 446–465.

[21] Janice Cuny and William Aspray. 2001. *Recruitment and retention of women graduate students in computer science and engineering: Report of a workshop, June 20-21, 2000.* Computing Research Association.

[22] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the conference on computer supported cooperative work*. 1277–1286.

[23] Reed Milewicz Denae Ford and Alexander Serebrenik. 2019. How remote work can foster a more inclusive environment for transgender developers. In *2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)*. IEEE, 9–12.

[24] W DuBow and AS Pruitt. 2018. NCWIT Scorecard: The Status of Women in Technology. *NCWIT, Boulder, CO* (2018).

[25] Nicolas Ducheneaut. 2005. Socialization in an open source software community: A socio-technical analysis. *Proceedings of the Conference on Computer Supported Cooperative Work* 14, 4 (2005), 323–368.

[26] Christina Dunbar-Hester. 2019. *Hacking Diversity: The Politics of Inclusion in Open Technology Cultures*. Vol. 21. Princeton University Press.

[27] Lillian T Eby, Tammy D Allen, Sarah C Evans, Thomas Ng, and David L DuBois. 2008. Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of vocational behavior* 72, 2 (2008), 254–267.

[28] Fabian Fagerholm, Alejandro S Guinea, Jürgen Münch, and Jay Borenstein. 2014. The role of mentoring and project characteristics for onboarding in open source software projects. In *Proceedings of the International Symposium on Empirical Software Engineering and Measurement*. 1–10.

[29] Yulin Fang and Derrick Neufeld. 2009. Understanding sustained participation in open source software projects. *Journal of Management Information Systems* 25, 4 (2009), 9–50.

[30] Samer Faraj and Lee Sproull. 2000. Coordinating expertise in software development teams. *Management science* 46, 12 (2000), 1554–1568.

[31] Roy T Fielding. 1999. Shared leadership in the Apache project. *Commun. ACM* 42, 4 (1999), 42–43.

[32] Klint Finley. 2017. Diversity in open source is even worse than in tech overall. *Wired Magazine Website* (2017).

[33] Ronald Aylmer Fisher. 1992. Statistical methods for research workers. In *Breakthroughs in statistics*. Springer, 66–70.

[34] Brian Fitzgerald. 2006. The transformation of open source software. *MIS quarterly* (2006), 587–598.

[35] Stacy Branham Foad Hamidi, Morgan Klaus Scheuer. 2018. Gender recognition or gender reductionism? The social implications of automatic gender recognition systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018)*. ACM, paper 8, pages 1–13.

[36] Denae Ford, Justin Smith, Philip J Guo, and Chris Parnin. 2016. Paradise unplugged: Identifying barriers for female participation on stack overflow. In *Proceedings of the Joint Meeting on Foundations of Software Engineering*. 846–857.

[37] Lustig K. Banks J. Parnin C. Ford, D. 2018. We Don't Do That Here" How Collaborative Editing with Mentors Improves Engagement in Social QA Communities. In *Proceedings of the 2018 ACM conference on human factors in computing systems (CHI 2018)*. Association for Computing Machinery, New York, NY, USA, 1–12.

[38] S. Spaeth G. Von Krogh, S. Haefliger and M. W. Wallin. 2012. Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly* 36 (2012), 649–676. Issue 2.

[39] R Stuart Geiger. 2017. Summary analysis of the 2017 github open source survey. *arXiv preprint arXiv:1706.02777* (2017).

[40] Jaco Geldenhuys. 2010. Finding the core developers. In *Proceedings of the Conference on Software Engineering and Advanced Applications*. IEEE, 447–450.

[41] Mohammad Gharehyazie, Daryl Posnett, Bogdan Vasilescu, and Vladimir Filkov. 2015. Developer initiation and social interactions in OSS: A case study of the Apache Software Foundation. *Empirical Software Engineering* 20, 5 (2015), 1318–1353.

[42] Tim Guilford and Marian Stamp Dawkins. 2005. Collaboration and creativity: The small world problem. *American journal of sociology* 111, 2 (2005), 447–504.

[43] Roger Guimera, Brian Uzzi, Jarrett Spiro, and Luis A Nunes Amaral. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 5722 (2005), 697–702.

[44] David Harrison, Kenneth Price, and Myrtle Bell. 1998. Beyond Relational Demography: Time and the Effects of Surface- and Deep-Level Diversity on Work Group Cohesion. *The Academy of Management Journal* 41, 1 (1998), 96–107.

[45] Israel Herraiz, Gregorio Robles, Juan JosÉ Amor, Teófilo Romera, and Jesús M. González Barahona. 2006. The Processes of Joining in Global Distributed Software Projects. In *Proceedings of the 2006 International Workshop on Global Software Development for the Practitioner* (Shanghai, China) *(GSD '06)*. Association for Computing Machinery, New York, NY, USA, 27–33. https://doi.org/10.1145/1138506.1138513

[46] Guide Hertel, Sven Niedner, and Stephanie Herrmann. 2003. Motivation of Software Developers in Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel. *Research Policy* 32, 7 (2003), 1159–77.

[47] Oosterbeek H. Van Praag M. Hoogendoorn, S. 2013. The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management science* 59 (2013), 1514–1528. Issue 7.

[48] James Howison and James D. Herbsleb. 2011. Scientific software production: incentives and collaboration. (2011), 513–522.

[49] Beth K Humberd and Elizabeth D Rouse. 2016. Seeing you in me and me in you: Personal identification in the phases of mentoring relationships. *Academy of Management Review* 41, 3 (2016), 435–455.

[50] Nasif Imtiaz, Justin Middleton, Joymallya Chakraborty, Neill Robson, Gina Bai, and Emerson Murphy-Hill. 2019. Investigating the Effects of Gender Bias on GitHub. (2019).

[51] Daniel Izquierdo, Nicole Huesman, Alexander Serebrenik, and Gregorio Robles. 2018. Openstack gender diversity report. *IEEE Software* 36, 1 (2018), 28–33.

[52] Chris Jensen and Walt Scacchi. 2007. Role migration and advancement processes in OSSD projects: A comparative case study. In *Proceedings of the International Conference on Software Engineering*. IEEE, 364–374.

[53] Corey Jergensen, Anita Sarma, and Patrick Wagstrom. 2011. The onion patch: migration in open source ecosystems. In *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering*. 70–80.

[54] Sandra Katz, David Allbritton, John Aronis, Christine Wilson, and Mary Lou Soffa. 2006. Gender, achievement, and persistence in an undergraduate computer science program. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 37, 4 (2006), 42–57.

[55] Bruce Kogut and Anca Metiu. 2001. Open-source software development and distributed innovation. *Oxford review of economic policy* 17, 2 (2001), 248–264.

[56] Kathy E Kram. 1985. Mentoring at Work: Developmental Relationships in Organizational Life (Scott, Foresman, Glenview, IL). (1985).

[57] Victor Kuechler, Claire Gilbertson, and Carlos Jensen. 2012. Gender differences in early free and open source software joining process. In *International Conference on Open Source Systems*. Springer, 78–93.

[58] Adriaan Labuschagne and Reid Holmes. 2015. Do onboarding programs work?. In *Proceedings of the Working Conference on Mining Software Repositories*. IEEE, 381–385.

[59] Karim R Lakhani and Robert G Wolf. 2003. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. (2003).

[60] Melenie J Lankau and Terri A Scandura. 2002. An investigation of personal learning in mentoring relationships: Content, antecedents, and consequences. *Academy of management Journal* 45, 4 (2002), 779–790.

[61] Andrew M St Laurent. 2004. *Understanding open source and free software licensing: guide to navigating licensing issues in existing & new software.* " O'Reilly Media, Inc.".

[62] Amanda Lee and Jeffrey C Carver. 2019. FLOSS participants' perceptions about gender and inclusiveness: a survey. In *Proceedings of the International Conference on Software Engineering*. IEEE, 677–687.

[63] John Levine and Moreland. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 5722 (2005), 697–702.

[64] Fannie Liu, Denae Ford, Chris Parnin, and Laura Dabbish. 2017. Selfies as social movements: Influences on participation and perceived impact on stereotypes. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–21.

[65] Sam A Margolis and Jacob B Angelo. 2002. Interlaboratory assessment of measurement precision and bias in the coulometric Karl Fischer determination of water. *Analytical and bioanalytical chemistry* 374, 3 (2002), 505–512.

[66] Christopher Mendez, Hema Susmita Padala, Zoe Steine-Hanson, Claudia Hilderbrand, Amber Horvath, Charles Hill, Logan Simpson, Nupoor Patil, Anita Sarma, and Margaret Burnett. 2018. Open source barriers to entry, revisited: A sociotechnical perspective. In *Proceedings of the International Conference on Software Engineering*. 1004–1015.

[67] Amanda Menking and Ingrid Erickson. [n.d.]. The Heart Work of Wikipedia: Gendered, Emotional Labor in the World's Largest Online Encyclopedia. ([n. d.]), 207–210.

[68] Courtney Miller, David Gray Widder, Christian Kästner, and Bogdan Vasilescu. 2019. Why do people give up flossing? A study of contributor disengagement in open source. In *Proceedings of the International Conference on Open Source Systems*. Springer, 116–129.

[69] Audris Mockus, Roy T Fielding, and James Herbsleb. 2000. A case study of open source software development: the Apache server. In *Proceedings of the international conference on Software engineering*. 263–272.

[70] Audris Mockus, Roy T Fielding, and James D Herbsleb. 2002. Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 11, 3 (2002), 309–346.

[71] Dawn Nafus. 2012. 'Patches don't have gender': What is not open in open source software. *New Media & Society* 14, 4 (2012), 669–683.

[72] Wonseok Oh and Sangyong Jeon. 2007. Membership herding and network stability in the open source community: The Ising perspective. *Management science* 53, 7 (2007), 1086–1101.

[73] Gede Artha Azriadi Prana, Denae Ford, Ayushi Rastogi, David Lo, Rahul Purandare, and Nachiappan Nagappan. 2021. Including everyone, everywhere: Understanding opportunities and challenges of geographic gender-inclusion in oss. *IEEE Transactions on Software Engineering* (2021).

[74] Sebastian Pölsterl. [n.d.]. Overview of Python Packages for Scientific Computing. https://gist.github.com/sebp/58da862b779489998e8e6088908fbfa5, Last accessed on 2021-04-15.

[75] Huilian Sophie Qiu, Yucen Lily Li, Susmita Padala, Anita Sarma, and Bogdan Vasilescu. 2019. The Signals that Potential Contributors Look for When Choosing Open-source Projects. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29.

[76] Huilian Sophie Qiu, Alexander Nolte, Anita Brown, Alexander Serebrenik, and Bogdan Vasilescu. 2019. Going farther together: The impact of social capital on sustained participation in open source. In *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 688–699.

[77] Lilly Irani Richard Geiger, Dorothy Howard. 2021. The Labor of Maintaining and Scaling Free and Open-Source Software Projects. *Proceedings of the ACM Human-Computer Interaction* 5 (2021), Article 175. Issue CSCW1.

[78] Q.M. Roberson. 2019. Diversity in the Workplace: A Review, Synthesis, and Future Research Agenda. *Annual Review of Organizational Psychology and Organizational Behavior* 6 (2019), 69–88.

[79] Jeffrey A Roberts, Il-Horn Hann, and Sandra A Slaughter. 2006. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management science* 52, 7 (2006), 984–999.

[80] Gregorio Robles, Jesus M Gonzalez-Barahona, and Israel Herraiz. 2009. Evolution of the core team of developers in libre software projects. In *Proceedings of the international working conference on mining software repositories*. IEEE, 167–170.

[81] Gerald R Salancik and Jeffrey Pfeffer. 1978. A social information processing approach to job attitudes and task design. *Administrative science quarterly* (1978), 224–253.

[82] Spiel Katta Haimson Oliver Hamidi Foad Branham Stacy M. Scheuerman, Morgan Klaus. 2020. HCI Guidelines for Gender Equity and Inclusivity. https://www.morgan-klaus.com/gender-guidelines.html, Last accessed on 2021-04-15.

[83] Kusum Singh, Katherine R Allen, Rebecca Scheckler, and Lisa Darlington. 2007. Women in computer-related majors: A critical synthesis of research and theory from 1994 to 2005. *Review of Educational Research* 77, 4 (2007), 500–533.

[84] Vandana Singh and Brice Bongiovanni. 2021. Motivated and Capable but No Space for Error. *The International Journal of Information, Diversity, & Inclusion (IJIDI)* 5, 3 (2021).

[85] Vandana Singh and William Brandon. 2019. Open Source Software Community Inclusion Initiatives to Support Women Participation. In *IFIP International Conference on Open Source Systems*. Springer, 68–79.

[86] Vibha Singhal Sinha, Senthil Mani, and Saurabh Sinha. 2011. Entering the circle of trust: developer initiation as committers in open-source projects. In *Proceedings of the Working Conference on Mining Software Repositories*. 133–142.

[87] Peijian Song and Chee Wei Phang. 2015. Promoting continuance through shaping members' social identity in knowledge-based versus support/advocacy virtual communities. *Transactions on Engineering Management* 63, 1 (2015), 16–26.

[88] Christoph Stanik, Lloyd Montgomery, Daniel Martens, Davide Fucci, and Walid Maalej. 2018. A simple nlp-based approach to support onboarding and retention in open source communities. In *Proceedings of the International Conference on Software Maintenance and Evolution*. IEEE, 172–182.

[89] Igor Steinmacher, Ana Paula Chaves, Tayana Uchoa Conte, and Marco Aurelio Gerosa. 2014. Preliminary empirical identification of barriers faced by newcomers to Open Source Software projects. In *Proceedings of the Brazilian Symposium on Software Engineering*. IEEE, 51–60.

[90] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. 2015. Social barriers faced by newcomers placing their first contribution in open source software projects. In *Proceedings of the Conference on Computer Supported Cooperative Work*. 1379–1392.

[91] Igor Steinmacher, Tayana Uchoa Conte, Christoph Treude, and Marco Aurélio Gerosa. 2016. Overcoming open source project entry barriers with a portal for newcomers. In *Proceedings of the International Conference on Software Engineering*. 273–284.

[92] Igor Steinmacher, Marco Aurelio Graciotto Silva, Marco Aurelio Gerosa, and David F Redmiles. 2015. A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology* 59 (2015), 67–85.

[93] Katherine J Stewart and Sanjay Gosain. 2006. The impact of ideology on effectiveness in open source software development teams. *Management Information Systems Quarterly* (2006), 291–314.

[94] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. *Handbook of qualitative research* 17 (1994), 273–85.

[95] Yuri Takhteyev and Andrew Hilts. 2010. Investigating the geography of open source software through github.

[96] Yong Tan, Vijay Mookerjee, and Param Singh. 2007. Social capital, structural holes and team composition: Collaborative networks of the open source software community. *ICIS 2007 Proc* (2007), 155.

[97] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson Murphy-Hill, Chris Parnin, and Jon Stallings. 2017. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer*

*Science* 3 (2017), e111.

[98] Chaihirunkarn C. Kalyanasundaram A. Trainer, Erik and James D. Herbsleb. [n.d.]. From personal tool to community resource: What's the extra work and who will do it? ([n. d.]), 417–430.

[99] Chaihirunkarn C. Kalyanasundaram A. Trainer, Erik and James D. Herbsleb. [n.d.]. Motivation, governance, and the viability of hybrid forms in open source software development. *Management science* 52, 7 ([n. d.]), 1000–1014.

[100] Bianca Trinkenreich, Igor Wiese, Anita Sarma, Marco Gerosa, and Igor Steinmacher. 2021. Women's Participation in Open Source Software: A Survey of the Literature. *arXiv preprint arXiv:2105.08777* (2021).

[101] Mariam Guizani Igor Wiese Anita Sarma Trinkenreich, Bianca and Igor Steinmacher. 2020. Hidden Figures: Roles and Pathways of Successful OSS Contributors. *Proceedings of the ACM Human-Computer Interaction* 4 (2020), 1–22. Issue CSCW2.

[102] Mariam Guizani Igor Wiese Anita Sarma Trinkenreich, Bianca and Igor Steinmacher. 2021. The shifting sands of motivation: Revisiting what drives contributors in open source. *arXiv preprint* arXiv:2101.10291 (2021).

[103] Asher Trockman, Shurui Zhou, Christian Kästner, and Bogdan Vasilescu. 2018. Adding sparkle to social coding: an empirical study of repository badges in the npm ecosystem. In *Proceedings of the International Conference on Software Engineering*. 511–522.

[104] Marat Valiev, Bogdan Vasilescu, and James Herbsleb. 2018. Ecosystem-level determinants of sustained activity in open-source projects: A case study of the PyPI ecosystem. In *Proceedings of the Joint Meeting on Foundations of Software Engineering*. 644–655.

[105] Bogdan Vasilescu, Vladimir Filkov, and Alexander Serebrenik. 2015. Perceptions of diversity on git hub: A user survey. In *Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering*. IEEE, 50–56.

[106] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *Proceedings of the ACM conference on human factors in computing systems*. 3789–3798.

[107] Georg Von Krogh, Sebastian Spaeth, and Karim R Lakhani. 2003. Community, joining, and specialization in open source software innovation: a case study. *Research policy* 32, 7 (2003), 1217–1241.

[108] Jianguo Wang and Anita Sarma. 2011. Which bug should I fix: helping new developers onboard a new project. In *Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering*. 76–79.

[109] Yi Wang and David Redmiles. 2019. Implicit gender biases in professional software development: An empirical study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 1–10.

[110] Zhendong Wang, Yi Wang, and David Redmiles. 2018. Competence-confidence gap: A threat to female developers' contribution on github. In *Proceedings of the International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 81–90.

[111] Joan C Williams and Rachel Dempsey. 2018. *What works for women at work: Four patterns working women need to know*. NYU Press.

[112] P. Wurzelova, F. Palomba, and A. Bacchelli. 2019. Characterizing Women (Not) Contributing to Open-Source. In *2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)*. IEEE.

[113] Bo Xu and Donald R Jones. 2010. Volunteers' participation in open source software development: a study from the social-relational perspective. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 41, 3 (2010), 69–84.

[114] Kazuhiro Yamashita, Shane McIntosh, Yasutaka Kamei, Ahmed E Hassan, and Naoyasu Ubayashi. 2015. Revisiting the applicability of the pareto principle to core development teams in open source software projects. In *Proceedings of the International Workshop on Principles of Software Evolution*. 46–55.

[115] Yue Yu, Huaimin Wang, Gang Yin, and Tao Wang. 2016. Reviewer recommendation for pull-requests in github: What can we learn from code review and bug assignment? *Information and Software Technology* 74 (2016), 204–218.

[116] Haiyi Zhu, Amy Zhang, Jiping He, Robert E Kraut, and Aniket Kittur. 2013. Effects of peer feedback on contribution: a field experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2253–2262.

[117] Stuart H Zweben and Elizabeth B Bizot. 2016. Representation of women in postsecondary computing: Disciplinary, institutional, and individual characteristics. *Computing in Science & Engineering* 18, 2 (2016), 40–56.

# A CODING SCHEME

## A.1 Initial involvement motivations

This section presents the codes related to participants' initial motivations for engaging with the open source project they discussed in Table 2.

Table 2. Codebook of motivations for initial project involvement

| Category | Code | Definition |
|---|---|---|
| Personal | Self-development | Joining or creating a project as a venue to learn a new skill or improve on an existing one. |
| | OSS Values | Involvement because they believe in sharing code, knowledge, openness, and Actions or behaviors that promote open source software ideals to a broader audience |
| | Social good | Creating or joining a project to support a specific group of people who would use or benefit from it |
| | Technical need | Creating or working on an existing open source project that addresses a technical barrier in a personal project carried out for fun, as a hobby or otherwise on a volunteer or unpaid basis. |
| | Personal interest | Creating or joining a project because of a specific passion, hobby, past experience or function they're interested in. |
| | Invitation | Joining a project because of direct invitation from another person. |
| Job-related | Job assignment | Assigned to work on open source project as part of their ongoing work tasks. |
| | Used for technical need | Selecting a specific project to incorporate into their work because it provides a potential solution to a work-related technical problem. |
| | Created for technical need | Creating a new open source project because they had solved a new issue related to work and decided to share their solution publicly (i.e created a license so anyone can use it) |

## A.2 Motivations for increasing involvement

This section presents the codes related to motivations for increasing involvement in the open source project in which the participant was most involved in Table 3.

Table 3. Codebook of motivations for increased project involvement

| Category | Code | Subcodes | Definition |
|---|---|---|---|
| Serving Users | User empathy | | Developers putting themselves in the user's shoes, or somehow trying to get a better understanding of the user's experience, which then influences their role and decisions in the project. |
| | Soliciting user feedback | | Seeking out and interacting with users on proposed changes or recent decisions to get the user's perspective, which may then influence future decisions about the project |
| | Responding to user requests | | Having direct interactions with users as a motivating factor, helping users solve a problem or get what they need, responding to user questions, issues, messages on Slack etc. |
| | Building relationships with users | | Long term relationships with specific user groups, eg clients. Positive affinity for specific users developed through repeated positive interactions / collaborative working arrangement. |
| Addressing project needs | Project management | Setting project direction / vision | Developer firmly believes in the vision and future of the project and becomes more involved to make that vision a reality |
| | | Internal work coordination and management | Developer wants to create a more streamlined experience for the core team on making decisions, organizing events, distributing tasks fairly, etc. |
| | Code and contribution management | Writing code to address perceived technical need | Developer wants to address a perceived flaw or shortcoming in the project and works harder to mitigate the issue |
| | | Reviewing code contributions | Becomes more involved with community and project state by reviewing issues, communicating with contributors, and discussing changes with core team |
| | Developing social infrastructure | Creating docs to enable participation | Being aware of how the project appears to those with less experience, they create more documentation to alleviate any entry barriers related to docs |
| | | Project infrastructure / community tools | Creating or setting up tools or technical infrastructure needed to communicate or collaborate with other developers on the project |
| | | Community building and leadership | Strengthening the feeling of community among project members and directly working with or bringing up new members or contributors to help them contribute to the project |
| | | Project advocacy / publicity | Publicly promoting the project on social media or through talks or appearances at conferences and workshops |
| Social factors | Community | | Feelings of belonging and connection with the community of project contributors |
| | Building relationships | | Relationship formation with other members. More interactions across more channels |
| | Signals work is valued | Increased access rights, Code used by others, Contribution accepted | Actions that indicate contributions are valued including contribution acceptance, increased access rights, code that the participant has written is being used by other developers as foundation for their work, by project users. |