UNDERSTANDING AND SUPPORTING WOMEN'S PARTICIPATION IN OPEN SOURCE SOFTWARE

By Bianca Trinkenreich

A Dissertation

Submitted in Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Informatics & Computing

Northern Arizona University

December 2022

Approved:

Igor Steinmacher, Ph.D., Chair

Marco Aurelio Gerosa, Ph.D.

Lisa Hardy, Ph.D.

Anita Sarma, Ph.D.

Daniel German, Ph.D.

ABSTRACT

UNDERSTANDING AND SUPPORTING WOMEN'S PARTICIPATION IN OPEN SOURCE SOFTWARE

BIANCA TRINKENREICH

Women represent less than 24% of employees in the software development industry and experience various types of prejudice and bias. Despite various efforts to increase diversity and multi-gendered participation, women are even more underrepresented in Open Source Software (OSS) projects. Many OSS communities are aware of the importance of correcting this imbalance and spend significant resources on a variety of onboarding and outreach programs targeted to women. However, these programs are insufficient if those who onboard do not feel that they belong to the community and ultimately abandon the project—a frequent occurrence in OSS. Previous research in psychology, health care, and education has shown that a sense of belonging is a basic human need that affects a broad variety of behaviors and has implications for long-term engagement and job satisfaction. Nevertheless, the investigation of a sense of belonging in OSS so far has been under-explored. It is still unclear what environmental factors contribute to a (lack of a) sense of belonging and how to improve it in practice. In my PhD, I investigate the following questions: How do women participate in OSS projects? How do different forces affect women's participation in a large and community-oriented OSS project? I worked with the Linux Kernel managers on a case study that has the ultimate goal to increase women's participation. Some problems surpass the organization and are related to the local culture of the OSS communities. There are problems that go beyond the company's gates and permeate society, which often contributes to this cultural legacy. However, there is also space for improvement. The results of this research include a theoretical framework

that describes open-source-specific factors that can impact women's participation in OSS projects.

ACKNOWLEDGEMENTS

This dissertation is the result of a long, exciting journey. I would like to take a moment to show my gratitude to those who are part of this achievement.

I am incredibly grateful to my advisor and co-advisor, Dr. Igor Steinmacher and Dr. Marco Aurelio Gerosa, respectively, for all the guidance, support, and encouragement over these years. Thank you for trusting my work and always being there for me inside and outside NAU. I also extend my gratitude to the committee members whose valuable contributions helped improve and shape this work. Thank you, Anita Sarma, who indefatigably guided and mentored me from the very beginning. Thank you, Lisa Hardy, for giving me the theoretical base I needed to consolidate my knowledge about qualitative research. And thank you, Daniel German, for picking me from a mailing list and inviting me to work with him and the Linux Kernel.

Completing my dissertation would not be possible without the support of Dr. Klaas Jan-Stol. He showed me the steps to learn Structural Equation Models in a smooth, pleasant, and gentle way. I would also thank Dr. Tayana Conte, who has graced me with special care and love since the beginning of my academic journey in 2015. I am deeply indebted to the Linux Kernel community managers Kate Stewart and Shuah Khan, for inviting me to lead a study about women in the Linux Kernel.

Thank you to my mom, dad, and sister, who supported my dream even while silently suffering from missing having us around. Finally, I do not have words to express my gratitude and love to the most important people in my life: Fabio and Mariana. Thank you for crossing the seas to live my dream; for suffering my pains. Thank you for understanding my absences and tiredness. I love you and I always will.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

To my daughter Mariana, who was born one month after I graduated from Master's degree and followed me for this long doctoral journey from three to six years old. To Fabio–a colleague and a husband, who inspires every step I take and is my best friend.

Chapter 1

INTRODUCTION

This Ph.D. dissertation investigates the different factors that influence women's participation in OSS projects. This chapter introduces the context and problem, research statement, goal and questions, research methodology, expected contributions, and publications.

1.1 Context and Problem

Open Source Software (OSS) development is a collaborative endeavor in which expert developers distributed around the globe create software solutions (Oreg and Nov, 2008; Forte and Lampe, 2013). OSS communities evolved from comprising a group of talented software hackers volunteering to produce high-quality software into a more mainstream and commercially viable model (Fitzgerald, 2006; Robles *et al.*, 2019; Steinmacher *et al.*, 2017). This new OSS landscape involves well-known companies that not only use OSS but also open-source their products, as well as both join and manage communities (Robles *et al.*, 2019). Steinmacher *et al.* (2017) recently referred to this as marking the end of OSS's teenage years.

In an era where OSS is ubiquitous and forms the digital backbone of our society (Eghbal, 2016), diversity in OSS projects is increasingly gaining attention. Diversity can take many different forms, including gender, experience, culture, technical knowledge, and cognitive thinking. Some teams are more diverse in one attribute and less in others (Vasilescu *et al.*, 2015b). Improving diversity is nowadays seen as a goal for fairness (Terrell *et al.*,

2017) and productivity (Vasilescu *et al.*, 2015b). For example, previous research shows that gender diversity positively affects productivity by bringing together different perspectives; improving outcomes (Vasilescu *et al.*, 2015b), innovation, and problem-solving capacity; and fostering a healthier work environment (Earley and Mosakowski, 2000).

Although organizations are taking action to increase gender diversity, the percentage of women in OSS projects is on average lower than 10%. Only 7.5% of the contributions to public code from the last 50 years were authored by women (Zacchiroli, 2020). Women represent only 5.2% of the contributors to the Apache Software Foundation (Sharan, 2016), 9.9% in Linux kernel (Bitergia, 2016), and 10% of OpenStack contributors (Izquierdo *et al.*, 2018), three of the largest and most well-known OSS communities. Indeed, women represent only 9% of GitHub users (Vasilescu *et al.*, 2015c). We use the term "gender" as a socially constructed concept (Butler, 1997) where gender identification, display, and performance might or might not align with the sex assigned at birth. To reflect this social concept of gender, we use the term "women" and "men" as a shorthand for people who identify as women or men, respectively. Our work can be extended to increase inclusiveness for all genders, but we focus on women, a highly underrepresented group in OSS. Considering the benefits of having a more gender-diverse team, researchers are also increasingly focusing on understanding the low representation of women in OSS.

The lack of diversity in OSS has been well-documented for years, and minimal progress has been made (Ford *et al.*, 2017; Robles *et al.*, 2016; Turkle, 2005a). Robles *et al.* (2016) refer to the lack of progress in diversity over the last 10 years as a "lost decade" in the inclusion of women in OSS. Research suggests that gender bias and sexist behavior pervade OSS (Nafus, 2012; Terrell *et al.*, 2017). Women feel frustrated when they are the only woman on a development team or when their input is under-valued or ignored, even

on topics in which they have expert knowledge (Vasilescu *et al.*, 2015a). Within OSS projects, the notion of meritocracy reigns, following the logic that quality speaks for itself and will be rewarded (Feller and Fitzgerald, 2000). Continually finding themselves on the bottom rung, it is no surprise that many women report experiencing "imposter syndrome" (Vasilescu *et al.*, 2015b). Gender biases can represent a persistent barrier for women to join OSS (Mendez *et al.*, 2018a,b).

Improving the state of gender diversity in OSS would require not only attracting but also retaining more women. While some of the factors that affect retention are hard to control—such as the popularity of the project, and how early in the project life-cycle a developer joins (Zhou and Mockus, 2014; Lin *et al.*, 2017)—there are also measures that communities can take to encourage retention. Promoting early interactions with peers can support contributors' retention (Zhou and Mockus, 2014) and increase the newcomers' sense of belonging (Dominic *et al.*, 2020). According to the literature, developers need to have a sense of belonging and social bond to feel connected to a group (Xu *et al.*, 2006). The sense of belonging in a group is key to being productive and feeling satisfied and engaged (Baumeister and Leary, 2017; Lin, 2008; Espinosa, 2011). Without such a sense of belonging, people might quit (turn over) (Dávila and García, 2012). Hagerty et al. (Hagerty *et al.*, 1992) define a sense of belonging as "the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment."

While motivations and challenges are known forces that can push women towards or against contributing to OSS (David and Shapiro, 2008; Gerosa *et al.*, 2021; Trinkenreich *et al.*, 2021c), no research uncovers perceptions of success or sense of virtual community in OSS. Moreover, no research shows the connection between motivations and sense of

virtual community in OSS.

1.2 Research Statement, Goals and Questions

Research Statement: A person's career decision-making can be influenced by many factors. Women's career choices are considerably more difficult to predict than men's (Creamer and Laughlin, 2005). Despite the growing need for workers in the tech industry, women are still underrepresented to take advantage of the career opportunities presented in the field. An important step in formulating strategies to encourage women to pursue IT careers is examining how the women in the field pursued their career pathways and the reasons why some women have chosen such a career path. An analysis of the influences and experiences that informed their decisions may help shape the policies and approaches taken by other women, educators, and industry leaders who wish to expand women's career choices. It is the purpose of this study to take this step by analyzing the motivations, career goals, challenges, and belonging to OSS projects that influenced the decision of women to participate in OSS projects.

Research Goals: In this dissertation, our goal is to explain the forces that push women towards or against OSS projects: the different career pathways, motivations to join and stay, perceptions of success, challenges and reasons to leave, and sense of virtual community. To achieve this goal, we designed two **research questions**:

- RQ1: How do women participate in OSS projects?
- RQ2: How do different forces affect women's participation in a large and community-oriented OSS project?

1.3 Research Plan

I organized my mixed methods research into five studies to accomplish its goal, as depicted in Figure 1.1. To answer RQ1 we conducted the first four studies (Chapter 3, 4, 5, and 6). In the fifth and sixth studies (Chapter 7), we focused in depth on the challenges and sense of virtual community in a specific OSS community, answering RQ2.

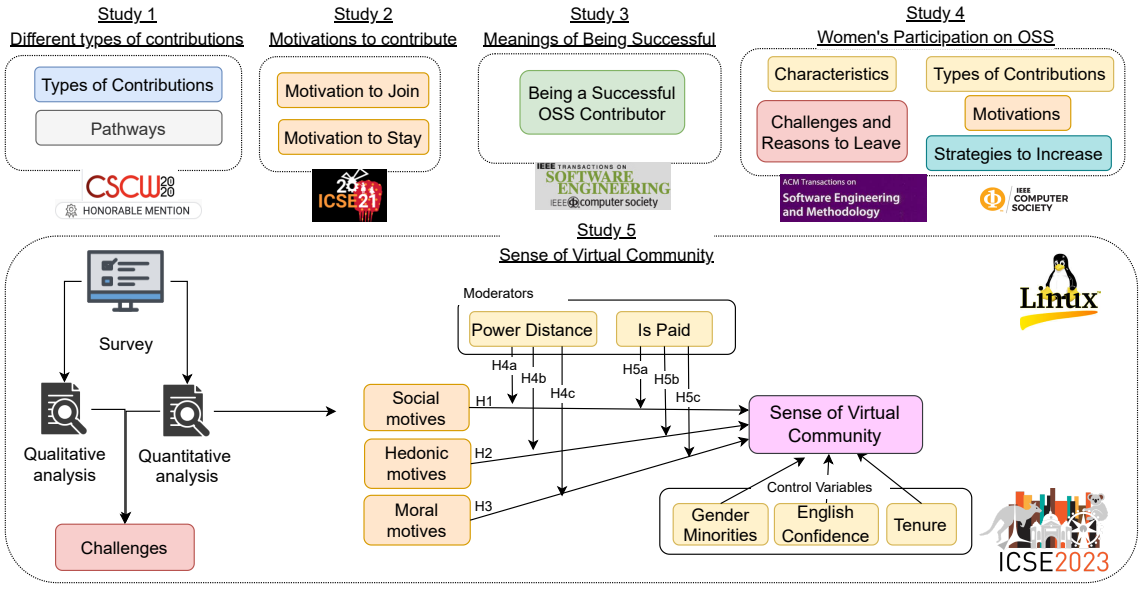

Figure 1.1: The Research Plan

We summarize the information flow between the studies in Figure 1.2 below. We present the studies in the order they were performed. The first, second, and third studies were based on inputs from contributors of different genders who make different types of contributions in different communities. The fourth study was based on existing literature. The fifth and sixth studies were performed in the Linux Kernel. Each study comprises a chapter with method and results:

- Career pathways and different types of OSS contributions. The first study we conducted was based on interviews, and the goal was to unveil the diversity of roles,

types of contributions, and pathways followed by OSS contributors (Trinkenreich *et al.*, 2020a). We present this study in Chapter 4.

- Motivations to contribute to OSS projects. The second study was based on an online survey and aimed to investigate how the MOTIVATIONS that drive contributors to participate in OSS projects shifted from 20 years ago, and how they shift after a contributor joins a project (Gerosa *et al.*, 2021). We present this study in Chapter 5

- Perceptions of success. The third study was based on both interviews and answers to an online survey, aiming to investigate, from the perspective of OSS contributors, MEANINGS OF SUCCESS (Trinkenreich *et al.*, 2021a). We present this study in Chapter 6.

- Systematic Mapping Study of Women's Participation in OSS. The fourth study comprised a systematic mapping of the literature about the women contributing to OSS: who they are, what motivates them to contribute, what types of contributions they make, the challenges they face, and the strategies that mitigate those challenges and increase women's participation in OSS (Trinkenreich *et al.*, 2021c, 2022b). We present this study in Chapter 3.

- Theoretical Model about Sense of Virtual Community. The sixth study developed a theoretical model of the intrinsic motivations that antecede the sense of virtual community for OSS contributors, and how culture and extrinsic factors interfere in this association. We used data from all genders to evaluate how language, tenure, and being part of a gender minority also affect the sense of virtual community (Trinkenreich *et al.*, 2023). We present this study in Chapter 7 - Section 7.2.

- The challenges faced by OSS contributors: The fifth study developed a framework of challenges reported by the Linux Kernel contributors. We used data from all genders and classified the challenges into four categories: process, technical, interpersonal, and personal. We present this study in Chapter 7 - Section 7.3.

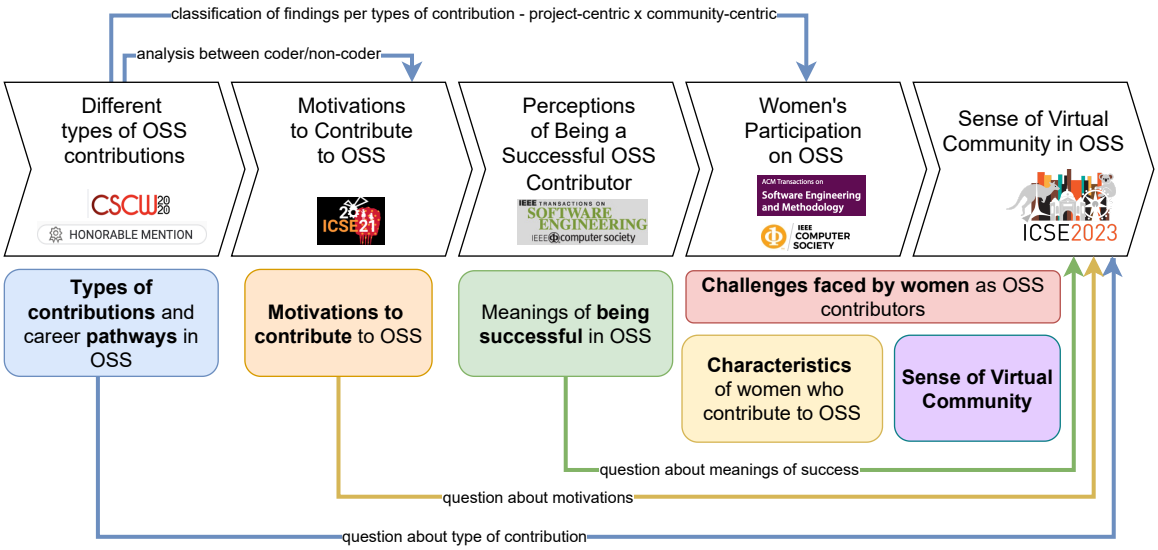

Figure 1.2: Research Flow and concepts explored per study

1.4 Contributions

Overall, this dissertation provides the following contributions:

- The multiple career pathways that can contributors can take, uncovering the hidden figures who play non-coding roles;
- The current motivations that drive contributors to join and stay in OSS;
- A multi-faceted definition of success for OSS contributors that sheds light on a diverse set of career goals;

- A theoretical model that links motivations and the sense of virtual community in OSS, and how this association is moderated by culture and payment, while being influenced by gender, English confidence, and tenure;

To the best of my knowledge, there is no work that provides a holistic view of forces that influence participation in OSS, including career pathways, motivations, perceptions of success, challenges, reasons to leave, and sense of virtual community.

1.4.1 Publications

The research conducted as part of this dissertation resulted in the following publications so far:

- Honorable Mention Award: TRINKENREICH, Bianca; GUIZANI, Mariam; WIESE, I., SARMA, A., STEINMACHER, I. 2020. Hidden figures: Roles and pathways of successful OSS contributors. In: *ACM Human-Computer Interaction, 4(CSCW2), pp.1-22* (Trinkenreich *et al.*, 2020a).

- TRINKENREICH, Bianca; GUIZANI, Mariam; WIESE, I.; CONTE, T.; GEROSA, M.; SARMA, A.; STEINMACHER, I. 2021. Pots of Gold at the End of the Rainbow: What is Success for Open Source Contributors? In: *IEEE Transactions of Software Engineering* (Trinkenreich *et al.*, 2021a).

- GEROSA, M.; WIESE, I.; TRINKENREICH, Bianca; STEINMACHER, I.; LINK, G.; ROBLES, G.; TREUDE, C.; SARMA, A.2021. The shifting sands of motivation: Revisiting what drives contributors in open source In: *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE) 2021 May 21 (pp. 1046-1058). IEEE* (Gerosa *et al.*, 2021).

- TRINKENREICH, Bianca; WIESE, I.; SARMA, A.; GEROSA, M.; STEINMACHER, I. 2021. Women's Participation in Open Source Software: A Survey of the Literature In: *ACM Transactions of Software Engineering and Methodology* (Trinkenreich *et al.*, 2021c)

- Best Paper Award: TRINKENREICH, Bianca; BRITTO, R.; GEROSA, M.; STEINMACHER, I. 2022. An Empirical Investigation on the Challenges Faced by Women in the Software Industry: A Case Study In: *2021 IEEE/ACM 44th International Conference on Software Engineering (ICSE) SEIS 2022 May 22* (Trinkenreich *et al.*, 2022a).

- TRINKENREICH, Bianca; GEROSA, M.; STEINMACHER, I. 2022. Women in Open Source: We Need to Talk About It In: *2022 IEEE Computer* (Trinkenreich *et al.*, 2022b).

- TRINKENREICH, Bianca; STOL, K.; GEROSA, M.; GERMAN, D.; SARMA, A.; STEINMACHER, I. 2023. Do I Belong? Modeling Sense of Virtual Community Among Linux Kernel Contributors In: *2022 IEEE/ACM 45th International Conference on Software Engineering (ICSE) Technical Track 2023 May 23* (Trinkenreich *et al.*, 2023).

I have also collaborated in research papers with researchers in OSS on diversity & inclusion (Guizani *et al.*, 2022) as well as about task recommendation for OSS contributors (Balali *et al.*, 2020), support for contributors to choose a task (Santos *et al.*, 2021, 2022), being an OSS Mentor Steinmacher *et al.* (2021), the challenges of Apache contributors Guizani *et al.* (2021).

1.4.2 Other Contributions

I served as a student volunteer for ICGSE 2020, ICSME 2020, and ICSE 2021. I also served as a Shadow Program Committee member for the 2022 Mining Software Repository conference, and on the Program Committee for ICSE, EASE, ASE, and OSS conferences. Additionally, I was a reviewer for several papers submitted to IEEE Transactions of Software Engineering (TSE), Software Quality Journal, IEEE Software, and Springer CSCW Journal. I accepted the invite to serve as Program Committee member for the Technical Track of 2023 Mining Software Repository conference.

Regarding awards, I was a finalist of the Student Research Competition awarded at ICSE 2021 and was awarded Outstanding Doctoral Student by the Northern Arizona University College of Engineering, Informatics, and Applied Sciences.

My research also had an impact on practice. This can be evidenced by the number of talks I gave at industry events. I was a speaker at the Linux Plumbers Conference (November 2021), where I talked about the motivations, career goals, challenges, and sense of virtual community of Linux Kernel contributors. I was also a speaker at the North America Open Source Summit (July 2022), where I talked about women's participation in Open Source, and at the Linux Foundation Member Summit (November 2022), about the different definitions of success in Open Source. This was previously presented to the Open Forum Europe at a forum regarding diversity and meritocracy in OSS (November 2021). Next, I was a speaker at Globant at the Women in Tech event (November 2022) and Tech Talks event (December 2022), where I talked about the impact of gender stereotyping and breaking the glass floor for women in the tech field.

1.5 Positionality of the Researcher

The concept of "positionality" is related to the dynamic and fluid ways a person is defined by either being part of an outsider or an insider dimension to the research (Maher and Tetreault, 2001; Louis *et al.*, 2002). Outsider researcher performs neutral and detached observations by not belonging to the group under investigation, whereas insider researchers study a group to which they belong and can authentically engage members of that group (Kerstetter, 2012). By sharing the group's experience, insider researchers can face the challenge of being critical and unbiased (Greene, 2014), whereas it can be hard for outsider researchers to fully understand something they have not experienced (Kerstetter, 2012). Previous research had offered a non-dichotomous definition for positionality, showing that instead of only insiders or outsiders, there are fluid nuances of positionality that allow us to reveal and unpack inherent power dynamics in research processes (Merriam *et al.*, 2001; Dwyer and Buckle, 2009).

I examine my positionality in terms of establishing transparency about my self attributes that were essential but potentially biased (Secules *et al.*, 2021). I identify my gender as a woman and aim to elevate women's voices across the software discipline, which is traditionally male-dominated and technically- and scientifically-demanding. As a woman in the software engineering space, I brought my experience to the research questions. Although I have worked in the software industry, I have never contributed to an Open Source Software project, and I consider myself an outsider in the investigated domain. During interviews, I disclosed my gender identity to help establish rapport and trustworthiness with interviewees. However, I was committed to suspending any connection of participant responses to my prior theoretical or personal knowledge.

Gender is one of the many diversity aspects that can impact the participation of an OSS contributor. Due to the small numbers of women (Section 1.1 and Section 3.1), we are not investigating (in this dissertation) the intersectionalities of gender and other structural differences (e.g. race, sexuality, economic status, politics). However, I understand those structural differences may make a difference and would like to continue exploring intersectionality in the future.

Chapter 2

BACKGROUND

This chapter presents an overview of: challenges faced by women in other domains and in STEM,; gender diversity in software engineering; and sense of virtual community in Section 2.3 and Section 2.4.

2.1 Challenges faced by women and strategies to increase women's participation outside OSS

2.1.1 Women in other domains

Analogous to OSS projects, women's barriers in the medical profession and their ability to rise to leadership positions are also influenced by social and cultural context (Ramakrishnan *et al.*, 2014). By contrast, women have played a significant active role in many contemporary armed rebellions (where men are often presumed to be the default gender) and even are frequently involved in leadership roles (Henshaw, 2016). By analyzing the career trajectories of women executives across a variety of sectors, Glass and Cook (Glass and Cook, 2016) concluded that while attaining promotion to leadership is not easy, serving in a high position can be even more challenging. Although women can be more likely than men to be empowered in high-risk leadership positions, they often lack the support or authority to accomplish their strategic goals. As a result, women leaders often experience shorter tenures compared to men peers (Glass and Cook, 2016). Similar to software development teams, where women are instrumental in reducing community smells Catolino

et al. (2019), in international relations the collaboration between women delegates and women civil society groups positively impacts and brings more durable peace when negotiating peace agreements (Krause *et al.*, 2018). The challenge of WORK-LIFE BALANCE that we present in Section 3.4 is a general challenge faced by women who aim to work in Japan, where the low numbers of women in medicine reflect the prevailing societal belief that careers and motherhood do not mix (Ramakrishnan *et al.*, 2014). In contrast, Scandinavia has similar numbers of men and women physicians, which has coincided with the emergence of progressive work-life policies, the belief that women can combine motherhood and employment, and changing expectations of work-life balance. Historically, Sweden was the first country to establish paid parental leave for fathers in 1974, and its National Labor Market Board has developed statements since 1977 encouraging men to contribute to childcare responsibilities (Haas and Hwang, 2009).

Kazmi (2014) and Maheshwari (2021) show that women in academia and industry, despite their impressive performance, face work-life balance issues and experience impostor syndrome (Kazmi, 2014), but can be supported through mentorship to overcome the challenges and advance in their career (Maheshwari, 2021).

More specifically for Computer Science, Pantic and Clarke-Midura (2019) found that individual (pre-arrival), institutional, and societal factors interplay on women's commitment and retention to the Computer Science program. Felizardo *et al.* (2021) found that research contributions from women in secondary studies have globally increased over the years, but are still concentrated in European countries. Although their findings are similar in terms of analyzing the factors that impact women's participation, none of them addressed specific motivations and challenges nor focused on strategies to increase women's participation in OSS.

2.1.2 Women in STEM

The literature shows that the gender gap is largely present in the science, technology, engineering, and mathematics (STEM) fields at all education levels and in the labor market (Fatourou *et al.*, 2019). The barriers faced by women in STEM fields, presented by McCullough (McCullough, 2011), are similar to some of the challenges we presented in Section 3.4, including discrimination and implicit bias (Toxic Culture), lifestyle choices, family obligations (Work-Life Balance Issues), and lack of role models and mentors (Community Reception Issues). Regarding strategies to increase women's participation, the W-STEM project (Garcia-Holgado *et al.*, 2019) seeks to create mechanisms to attract and guide women in Latin America in STEM higher education programs, including monitoring gender equality in enrollment and retention (as we present in the strategy Promote awareness of the presence of peers). Another strategy proposed by both Garcia-Holgado et al. (Garcia-Holgado *et al.*, 2019) and Moreno et al. (Moreno *et al.*, 2014) is to expose women to scientific and technological culture from an early age, promoting STEM studies vocation and choice to girls and young women in secondary schools (as we presented in the strategy Promote women-specific groups and events). Moreover, the women who participated on Moreno et al.'s (Moreno *et al.*, 2014)'s study corroborated the strategy of promote inclusive language and the need to avoid chauvinistic attitudes, and to de-stereotype the CS student, suggesting that the general image of the student majoring in CS should change and not ascribe to the nerd stereotype typically only ascribed to men.

2.2 Gender Diversity in Software Engineering

Gender (with a focus on women), is the most explored aspect of diversity in software engineering literature (Menezes and Prikladnicki, 2018; Silveira and Prikladnicki, 2019). Spichkova *et al.* (2017) analyzed the role of women within software architecture literature and found that it is understudied. In a more general literature review in Software Engineering, Rodríguez-Pérez *et al.* (2021) highlighted that researchers have been exploring gender bias problems in software engineering moreso than presenting solutions to mitigate these problems.

The systematic literature review (SLR) conducted by Dias Canedo *et al.* (2019) investigated the causes of women's lack of engagement in software development in general and found a list of challenges and possible solutions to increase women's engagement in OSS. Diversity in OSS has gained considerable attention in recent years, with OSS projects and foundations investing in efforts to create diverse and more inclusive communities. Research has also investigated the topic of low diversity and barriers to contributing to OSS. This is similar to software engineering research, a majority of which has focused on gender diversity, investigations of gender distribution in OSS (Bosu and Sultana, 2019; Robles *et al.*, 2016; Ortu *et al.*, 2017; Lin and Serebrenik, 2016; Gila *et al.*, 2014; Izquierdo *et al.*, 2018; Robles *et al.*, 2014) and in leadership positions Canedo *et al.* (2020), perceptions of women contributors in OSS Lee and Carver (2019); Vasilescu *et al.* (2015a), the impact of gender on productivity (Vasilescu *et al.*, 2015b), and the barriers that women face (Terrell *et al.*, 2017; Nafus, 2012; Mendez *et al.*, 2018a; Prana *et al.*, 2021; Wang *et al.*, 2018).

Vasilescu *et al.* (2015a) used a gender lens to understand GitHub contributors' perception of their team and awareness of their teammates' backgrounds, finding gender as the

second-most noticed attribute. Other research focused on women's experience in OSS and the support systems implemented to increase women's participation (Singh, 2019b; Singh and Brandon, 2019). Singh and Brandon (2019) found that only 12 out of 355 OSS websites have 'women only' sections and Lee and Carver (2019) found that while some contributors welcomed women's participation in OSS, some were strongly opposed to it. Finally, researchers have investigated barriers that women face in tools and technology (Mendez *et al.*, 2018a; Padala *et al.*, 2020), pull request acceptance (Terrell *et al.*, 2017), and inclusion in discussions (Nafus, 2012; Prana *et al.*, 2021).

Other diversity aspects are being investigated in OSS. The experience of "older" contributors in OSS (Murakami *et al.*, 2017; Morrison *et al.*, 2016; Davidson *et al.*, 2014) is explored to understand how age can impact code reviews. Murakami *et al.* (2017) found that age has no significant effect on code review correctness and efficiency. Morrison *et al.* (2016) investigated the low participation of veteran software developers in OSS and how their contributions differ from those of their younger peers. Morrison *et al.* (2016)'s results reported that veteran OSS contributors are less socially motivated than their younger counterparts, which aligns with Davidson *et al.* (2014) findings that older contributors face more social than technical challenges. Geolocation is also being examined to explore its impact on pull request acceptance. Furtado *et al.* (2020) found that contributors from countries with low human development indexes face the most pull request rejections. Similarly, Rastogi *et al.* (2018) investigated the top countries with the highest and lowest pull request acceptance rates and Rastogi *et al.* (2016) found pull request acceptance rate increases by 19% when the submitter and integrator are from the same country. Recent works have started to investigate diversity through the lens of multiple demographic attributes. For example, Prana *et al.* (2021) investigated the difference in gender diversity between geo-

graphic regions and found that there has been a small improvement in gender diversity amongst contributors in Northern America and South-Eastern Asia. Ortu *et al.* (2017) also used a dual-lens approach and found that gender diversity increased productivity, while intra-team nationality diversity decreased the level of politeness.

2.3 Sense of Belonging

A sense of belonging refers to the extent to which individuals feel like they belong or fit in a given environment (Sax *et al.*, 2018). Hagerty *et al.* (1992) posited that a sense of belonging represents a unique mental health concept that differs from concepts such as loneliness, alienation, and social support (Hagerty and Patusky, 1995). They define a sense of belonging as "the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment" (Hagerty *et al.*, 1992). Hagerty *et al.* (1992) delineated two defining attributes for belonging (Hagerty *et al.*, 1992): (i) valued involvement, or the experience of feeling valued, needed, or accepted; and (ii) fit, the perception that the individual's characteristics match with the system or environment. In this project, we explore these two components in the context of OSS communities.

The need to belong is a powerful, fundamental, and pervasive motivation with multiple and strong effects on emotional patterns and cognitive processes (Baumeister and Leary, 2017) and found across all cultures and types of people (Baumeister and Leary, 2017). Maslow positions belonging as a basic human need (Maslow, 1943). According to Baumeister and Leary (2017), a sense of belonging has affective consequences, elicits goal-oriented behavior, affects a broad variety of behaviors, and has implications that go beyond immediate psychological functioning. Belonging to a group enables a person to

feel valued (McDougall, 2015).

References to the importance of a sense of belonging are found throughout the psychological, health care, and education literature. A lack of a sense of belonging is linked to a variety of ill effects on health, adjustment, and well-being (Baumeister and Leary, 2017). Anant (Anant, 1966, 1967, 1969) reported an inverse relationship between sense of belonging and anxiety, and Lim (Lim, 2008) showed that sense of belonging is related to job satisfaction. Within the higher education literature, a sense of belonging is a known predictor of success in college (e.g., (Freeman *et al.*, 2007; Pittman and Richmond, 2008; Strayhorn, 2012)), especially within STEM disciplines Espinosa (2011); Johnson (2012). College students who feel a greater sense of belonging are more connected to their environment and are more likely to persist through to graduation (Strayhorn, 2018). Lacking a sense of belonging in college is associated with outcomes such as depression, anxiety, suicide, criminality, and college freshmen attrition (Hoyle and Crawford, 1994). Studies focused on computer science programs also highlight the importance of belonging in attracting and retaining students (Barker *et al.*, 2010; Cohoon, 2002; Margolis and Fisher, 2002; Veilleux *et al.*, 2012).

The literature specifically highlights the importance of belonging for underrepresented groups (Espinosa, 2011; Johnson, 2012; Thoman *et al.*, 2014). Studies focused on STEM students show that women have a lower sense of belonging than their counterparts and thus face additional obstacles and barriers to their success and retention in college (Beyer *et al.*, 2004; Cheryan *et al.*, 2009; Hurtado and Carter, 1997; Johnson *et al.*, 2007; Strayhorn, 2012). Espinosa (2011) found that a sense of belonging is significantly related to STEM retention. The literature reports that women experience computing environments differently due to sexism and racism, both historically and as part of the current culture (Barker *et al.*, 2009;

Margolis and Fisher, 2002; Margolis *et al.*, 2017; Strayhorn, 2012), potentially leading them to feel unwelcome and experience a lack of sense of belonging in computing spaces (Sax *et al.*, 2018). Although the literature has shown the importance of belonging to multiple contexts, to the best of our knowledge it has not been studied in OSS, which is plagued by low participation by women (Vasilescu *et al.*, 2015c; Zacchiroli, 2020; Kofink, 2015).

A sense of belonging can be influenced by a number of individual characteristics and factors of the surrounding environment (Allen, 2020). For example, in the education literature, several researchers (Anderman, 1999; Goodenow, 1993; Goodenow and Grady, 1993; Battistich *et al.*, 1995; Solomon *et al.*, 1996) found associations between students' sense of belonging and a range of motivational variables. Different perceptions of success and long term goals also were found to be associated with differences in feelings of belonging Anderman and Anderman (1999).

2.4 Sense of Virtual Community

While numerous definitions of the term 'community' exist, a common theme is that it involves human relationships based on some common characteristics Gusfield (1975). The classical McMillan and Chavis McMillan and Chavis (1986) definition of 'Sense of Community' includes four characteristics: (1) feelings of membership (belonging to, and identifying with, the community), (2) feelings of influence (having an influence on, and being influenced by the community), (3) integration and fulfillment of needs (being supported by others in the community while also supporting them), and (4) shared emotional connection (relationships, shared history, and a 'spirit' of community). *Virtual* communities typify a relatively new form of interaction, whereby community members share information and knowledge in the virtual space for mutual learning, collaboration, or

problem-solving Koh *et al.* (2003).

The development of OSS involves distributed problem-solving within a virtual community Martínez-Torres and Díaz-Fernández (2014). Virtual communities are a particularly important type of virtual group, because they are self-sustaining social systems in which members engage and connect with each other through their members' feelings of community, referred to as their sense of virtual community (SVC) Rheingold (2000). Sense of virtual community includes membership, identity, belonging, and attachment to a group that primarily interacts through electronic communication Blanchard (2007); Chang *et al.* (2016); Brown and Pehrson (2019). SVC tailors McMillan's theory of sense of community to the study of virtual communitiesMcMillan and Chavis (1986), with the goal to assess their "community-ness" Blanchard (2007).

SVC can be assessed and promoted by community managers to fulfill a core set of members' needs Sutanto *et al.* (2011) so they feel they belong to a unique group. Such meaningful relationships are associated with increased satisfaction and communication with the virtual community, trust Blanchard and Markus (2002), and social capital in the project the community is working at Zhao *et al.* (2012). SVC has been shown to lead to an occupational commitment Blanchard *et al.* (2011), and ultimately can help retain contributors and further attract potential newcomers Blanchard (2007); Chen *et al.* (2013).

SVC can be developed by *exchanging support* (Blanchard *et al.*, 2011; Tonteri *et al.*, 2011), *creating identities and making identifications* (Blanchard *et al.*, 2011), *producing mutual cognitive and affective trust* amongst members of a community (Blanchard *et al.*, 2011; Chih *et al.*, 2017; Lee, 2010), establishing norms, and a concertive control (Gibbs *et al.*, 2019).

2.5 Concluding remarks

Several forces influence women's decisions to start contributing and to remain in or leave an OSS project Steinmacher *et al.* (2014b). Women face barriers in several professions, and also in STEM, software engineering, and more specifically, OSS. Most of the challenges faced by women in OSS are social-based, including missing having other women around (lack of peer parity), encountering offensive language in mailing lists (non-inclusive communication), suffering symbolic violence and harassment (toxic culture), avoiding initiating a pull-request due to feeling unsafe and having lack of confidence (problems of impostor syndrome), facing challenges finding mentors (community reception issues), being boxed into specific roles (stereotyping), sharing time between work and family (work-life balance issues), and biases or lower acceptance rates of contributions when they explicitly identify themselves as women (gender-biased peer review) Trinkenreich *et al.* (2021c). In the next chapter, we report the related works about women in OSS. In the next chapter, we present a literature review focusing specifically on women in OSS projects, aiming to comprehensively understand the current state-of-the-art.

Chapter 3

RELATED WORK

This chapter presents existing literature about women's participation in OSS projects. To provide a snapshot of the state-of-the-art, we reviewed, summarized, and synthesized the current state of research on women's participation in OSS through a systematic mapping of the literature. The study included database search, backward and forward snowballing, and input from prolific authors; we selected and retrieved information from 51 primary studies published between 2000 and 2021. Our contributions emerged from our literature analysis, namely the characterization of women's participation, motivation, types of contributions, challenges, and the identification of strategies proposed in the literature to promote women's participation (Trinkenreich *et al.*, 2021c).

The research design for this study is summarized in Figure 3.1.

According to the literature, there are few women OSS contributors; most of them are recent contributors who make both non-code and code contributions. Only about 5% of projects were reported to have women as core developers. While women authors comprise less than 5% of pull requests, they have similar or higher rates of merge acceptance than men. Besides learning new skills and altruism, reciprocity and kinship are especially relevant motivations for women, who may leave a project if they are not appropriately compensated for their contributions. Women's challenges are mainly social, including a lack of peer parity and non-inclusive communication from a toxic culture. We found ten strategies reported in the literature, which were mapped to the reported challenges. The results from this study were accepted for publication in the Transactions on Soft-

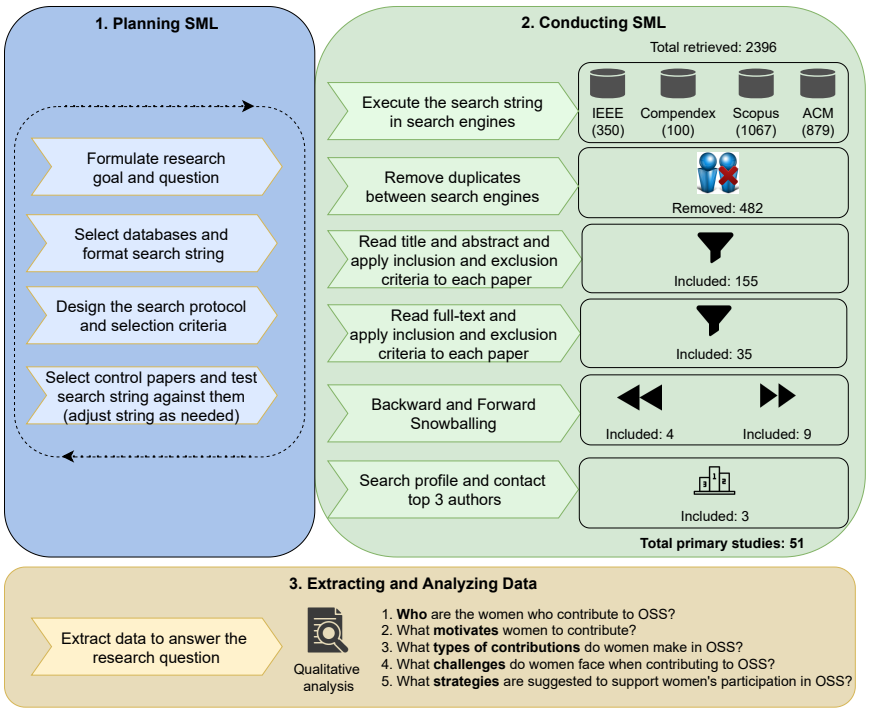

Figure 3.1: Research design for literature survey about women's participation in OSS

ware Engineering and Methodology (TOSEM) journal (Trinkenreich *et al.*, 2021c), and are discussed as follows.

3.1 Who are the women who contribute to OSS projects?

The literature presents different statistics for women's representation in OSS, which were measured in different ways, at different moments, and considering a different set of projects. We collected these results to provide a comprehensive view of the state of women's participation. Organizing data obtained from different sources may provide researchers with an expanded view of the phenomenon under study, prompting new insights and allowing for further discoveries (Robles *et al.*, 2014). Additionally, analyzing women's representation at different time periods can help us understand the evolution of gender imbalance. Understanding the characteristics (e.g., education level, family sta-

tus, diversification of projects, time to volunteer, tenure) can help communities and researchers to refine their strategies for attracting profiles of women who still do not participate and retain those who currently contribute to OSS.

Several primary studies quantified women's participation using a variety of methods, including mining software repositories and mailing lists, surveys, and participation in mentorship programs.

Participation of women in OSS found via mining software repositories. From a dataset of 23,493 GitHub projects, Vasilescu *et al.* (2015c) used the genderComputer tool (Vasilescu *et al.*, 2014) (with 93% of precision) to identify the gender, based on personal names, and, if available, countries, of 873,392 GitHub contributors. They found 91% men and 9% women. From the 5,250 OpenStack contributors, Izquierdo *et al.* (2018) inferred the gender using the genderize.io tool and found that 10% are women. From a dataset of 8,338 GitHub projects, Prana *et al.* (2021) found that the percentage of new GitHub accounts created by women has remained around 10% between 2014-2018. Bosu and Sultana (2019) analyzed a dataset of 683,865 code review requests from 10 popular OSS projects. Authors inferred gender using Gerrit-Miner tool (Bosu and Carver, 2013) and found that women represent 6.70% (out of 4,543) of non-casual developers (those who submitted at least five code changes) and only 4.27% (out of 936) of core developers (those who are the top 10% developers in terms of the number of code commits in a project). From a random sample of 300,000 GitHub users from a dataset with 16M users, Qiu *et al.* (2019b) inferred the gender using genderComputer and NamSor tools and identified 9.7% as women. Terrell *et al.* (2017) analyzed a GitHub dataset with 4,037,953 profiles and identified the gender of 1,426,127 (35.3%) through their public Google+ profiles. From those profiles, the authors analyzed pull-request submission and acceptance by women and men and found that 8,216

of the pull-requests were submitted by women (5.2%) and 150,248 (94.8%) by men. Imtiaz *et al.* (2019) used the same GitHub dataset of Terrell *et al.* (2017) and identified 529,253 men (93.7%) and 35,676 women (6.3%). Kofink (2015) also analyzed a dataset of 1,811,631 pull-requests and found that 4.5% were submitted by women and 95.5% by men. Zacchiroli (2020) also analyzed the authors of contributions. With 1.6 billion commits from the combined projects of GitHub, GitLab, and other development forges (using the Software Heritage project [1]), corresponding to the development history of 120 million projects, the author found that contributions were authored by 33 million distinct people over 50 years, and that there is significant growth of active women authors from around 4% in 2005 to 10% in 2019.

Distribution of women in OSS found by mining mailing lists: Kuechler *et al.* (2012) analyzed participation in eleven mailing lists of six projects (Buildroot, Busybox, Jaws, Parrot, uClibc, and Yum), which totaled 3,310 subscriptions. Authors found low participation by women: 8.27% of all subscribers, 6.63% of those who posted one message, 2.5% of those who posted more than ten times, and 1.5% of code reporters. Vasilescu *et al.* (2014) also used the mailing lists of two projects (Drupal and Wordpress) to explore women's representation and found that women authored 9.81% of the messages in Drupal and 7.81% in Wordpress. In contrast, both men and women engage in OSS projects for statistically similar lengths of time.

Distribution of women who participate in mentorship programs: By analyzing the gender of Google Summer of Code participants from 2016 to 2018, Dias Canedo *et al.* (2019) found that while there is a minor variation across the years, the volume of women stayed close

[1] https://www.softwareheritage.org/

to 11.98% of the total number of participants in the program. [2]

Distribution of women in OSS through surveys: Mani and Mukherjee (2016) and Robles *et al.* (2016) analyzed the same OSS 2013 survey data (Arjona-Reina *et al.*, 2014). This survey was answered by 2,183 OSS contributors, 226 of whom identified as women (10.35%). Lee and Carver (2019) received 119 answers to their questionnaire, wherein 10.92% of respondents identified as women, while Gerosa *et al.* (2021)'s questionnaire received 224 answers with 7.6% who identified as women.

In summary, the primary studies reported women's participation ratios ranging between 4% to 14% across different measurements and OSS communities. When we analyze the distribution over time based on when the primary studies were published (Fig.3.2), barring some fluctuation, women's participation ratio stays stable at around 10%. However, when taking a broader view, Zacchiroli (2020)'s analysis of public code contributions over the last 50 years found that women's contributions appear to be on the rise and are rising faster than those by male authors. This shows that while much still needs to be done, OSS projects are getting more gender diverse, albeit slowly.

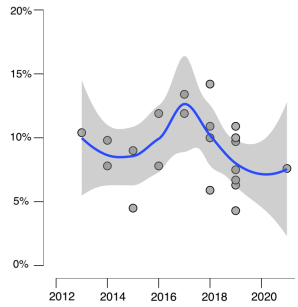

Figure 3.2: The frequency of women's participation in OSS as reported by primary studies per year. The values comprise percentages of women measured by different mechanisms.

[2]The Google Summer of Code (GSoC) is a 3-month OSS engagement program that offers stipends and mentorship to students as new contributors (Silva *et al.*, 2017; Trainer *et al.*, 2014; Silva *et al.*, 2020b).

To get a deeper understanding of the demographics of women who contribute to OSS projects, we analyzed available data reported in the primary studies along the following criteria: education level, time dedicated to contributions, diversification of projects, family status, and tenure (Fig 3.3).

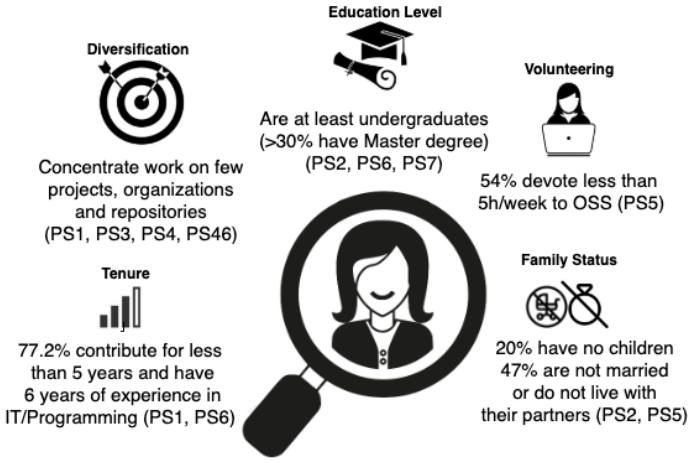

Figure 3.3: The characteristics of women who contribute to OSS projects

Education Level. Based on Stack Overflow's 2018 developer survey that included 43,000+ OSS contributors, Wurzelová *et al.* (2019) study reported that 82.8% of women who contribute to OSS are at least undergraduate students, compared to 76% of all contributors (Overflow, 2018). Mani and Mukherjee (2016) found a similar rate (81.4%) using data from the 2013 OSS Survey (226 women), while the corresponding figure for the whole dataset is 72%. All the 36 developers who identified as women on the survey from Canedo *et al.* (2020) were at least undergraduate students. A substantial number of contributors from the three studies were post-graduates, who either achieved a Master's degree—11.6% (Canedo *et al.*, 2020), 27.6% (Wurzelová *et al.*, 2019)— or Ph.D.—4.2% Wurzelová *et al.* (2019), 10% (Mani and Mukherjee, 2016), 22.9% (Canedo *et al.*, 2020).

Volunteering. Approximately half of the 226 women (53.59%) who answered the

2013 OSS questionnaire devote less than five hours per week to OSS projects. Only 14.77% of the women who answered this questionnaire dedicate more than 40 hours per week (Robles *et al.*, 2016), which can represent OSS as a full-time job. Although not mentioning the number of hours per week, according to Powell *et al.* (2010)'s results, 89% of women said they contribute to OSS both at home and at work, which includes bringing their work home and contributing to OSS during their leisure time.

Diversification of projects. By mining software repositories, three studies (Imtiaz *et al.*, 2019; Terrell *et al.*, 2017; Qiu *et al.*, 2019b) concluded that women concentrate their efforts on fewer projects than men. From a dataset comparing 152,534 pull-requests created by 20,926 women and 3,135,384 pull-requests created by 308,062 men, Imtiaz *et al.* (2019) concluded that women's pull-requests are concentrated in fewer projects and fewer organizations. Indeed, Qiu *et al.* (2019b) analyzed a balanced sample of the dataset, including 28,995 women and 29,096 men, and also concluded that women tend to concentrate their contributions on fewer different projects than men. From a dataset of 1,426,127 users whose gender could be identified, Terrell *et al.* (2017) analyzed the acceptance of submitted pull requests and concluded that women contribute to fewer projects than men. From another perspective, Vasilescu *et al.* (2015a) ran a survey answered by 199 women and 611 men, in line with the previous studies, and concluded that women own fewer public repositories than men.

Family Status. Almost half of the 226 women who took part in the OSS 2013 survey are not married or did not live with their partners (Mani and Mukherjee, 2016). This rate was composed of 35% of single women, over 11% women not living with their partners, 3% living with their partners, 3% married, 0.1% separated from their partners (Mani and Mukherjee, 2016); 20% of the women in that survey have children (Mani and Mukherjee,

2016; Robles *et al.*, 2016).

Tenure. Most (77.2%) of the 199 women who answered Vasilescu *et al.* (2015a)'s questionnaire have been contributing to open source projects for fewer than five years. Authors found that women have on average six years of experience in IT/programming, a significantly lower tenure than men, who have nine years of experience.

Women are underrepresented in central OSS roles, although they are better represented earlier in the joining process (e.g. in mentoring programs). The majority of women are recent contributors and can devote a few hours per week to OSS. Several studies found women's participation to range from 4.3% to 14.2%.

3.2 What motivates women to contribute to OSS projects?

Research has shown that women generally are more motivated to use technology to accomplish a goal rather than for fun (Burnett *et al.*, 2010). For the past 20 years, much academic work has theorized about and empirically examined OSS contributors' motivations. Retrieving and consolidating women's motivations from the existing studies is relevant to communities seeking to recruit and retain women. Proper management of motivation and satisfaction helps software organizations achieve higher productivity levels and avoid turnover, budget overflows, and delivery delays (Beecham *et al.*, 2008; França *et al.*, 2011; da Silva and França, 2012).

We consolidated the studies reporting women's motivation to participate in OSS projects, aggregating the results according to Von Krogh *et al.* (2012)'s categories. Von Krogh *et al.* (2012) surveyed the literature and identified ten categories of motivation, grouped as intrinsic, internalized-extrinsic, and extrinsic. Intrinsic motivation (enjoyment and fun, kinship, ideology, and altruism) moves a person to act for the fun or challenge entailed rather

than in response to external pressures or rewards (Ryan and Deci, 2000). In contrast, extrinsic motivations (Career and Pay) are based on outside incentives causing people to change their actions (Frey, 1997). Contributors can also internalize extrinsic motivators (learning, own-use, reciprocity, and reputation) as self-regulating behavior, such that it no longer needs to be externally imposed (Deci and Ryan, 1987; Roberts *et al.*, 2006). We summarize our findings organized in these higher-level categories in Figure 3.4.

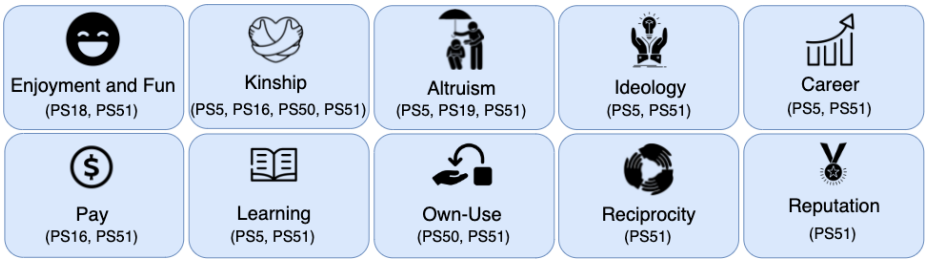

Figure 3.4: Women's motivations to contribute to OSS. Based on (Von Krogh *et al.*, 2012)

Among all 10 categories, the motivation literature showed that women see Enjoyment and Fun, Reciprocity, Kinship, and Pay somewhat differently from the motivation literature (Hertel *et al.*, 2003; Hars and Ou, 2004; Lakhani and Wolf, 2003; Ghosh *et al.*, 2002; Gerosa *et al.*, 2021), which is based on surveys predominantly answered by men. Enjoyment and Fun has been consistently reported as a top driver to contribute to OSS in motivation surveys (Lakhani and Wolf, 2003; Hars and Ou, 2004; Ghosh *et al.*, 2002; Hertel *et al.*, 2003; Gerosa *et al.*, 2021). However, according to the stratified analysis in Gerosa *et al.* (2021)'s study, no women reported this motivation to join (and one–6%–stayed because of this), while 7% of the men joined motivated by fun and 20% continued because of it.

The opposite trend is observed for Reciprocity, Kinship, and Pay. Reciprocity appeared as one of the top motivators for women in Gerosa *et al.* (2021)'s work—39% of

the women reported this as motivation to continue (versus 15% of the men). In general surveys (Lakhani and Wolf, 2003; Gerosa *et al.*, 2021), reciprocity is not among the top motivators. Regarding KINSHIP, while 39% of the 226 women who answered the OSS survey joined because of this motivation, 31% continued because of it (Robles *et al.*, 2016). Most (64%) of the 22 women who answered Prana *et al.* (2021)'s survey select a project in which friends and colleagues also participate. As part of kinship, peer parity also plays a role in women's motivation (David and Shapiro, 2008). David and Shapiro (2008) found that social connections with other developers influence women's choices. However, no women from Gerosa *et al.* (2021)'s study joined OSS because of kinship. Kinship is not top ranked in general surveys (Hars and Ou, 2004; Lakhani and Wolf, 2003; Ghosh *et al.*, 2002), but this trend has been changing with the rise of social coding platforms Gerosa *et al.* (2021). Regarding PAY, Prana *et al.* (2021) indicated that payment is a greater incentive for women than men (64% vs. 35%). This was echoed by women interviewees from Balali *et al.* (2018)'s study. However, the difference was not noticed in Gerosa *et al.* (2021)'s work, in which men and women equally reported money as a reason to continue contributing (14% and 11%, respectively).

For other motivations, we found that women follow a similar trend to that reported in the general literature. LEARNING, for example, frequently has been reported as a key motivation to contribute to OSS (Lakhani and Wolf, 2003; Hertel *et al.*, 2003; Hars and Ou, 2004; Ghosh *et al.*, 2002). Most (68%) of the 226 women who took part in the OSS 2013 survey joined for LEARNING and 65% continue because of it (Robles *et al.*, 2016). The same occurs for ALTRUISM, which is a common motivator in OSS (Gerosa *et al.*, 2021; Ghosh *et al.*, 2002) and relevant for women as well—37% of the women who took the OSS 2013 survey reported that they joined to share knowledge (Robles *et al.*, 2016) and 22% from

another work continued because of it (Gerosa *et al.*, 2021).

Similar rates were found for Career, Own-Use, and Reputation. Regarding Career, only 4% of the 226 women who took the OSS 2013 survey and none who took Gerosa *et al.* (2021)'s survey reported joining to improve their careers. Similar rates were found for men (5% and 8%, respectively). For Own-Use, David and Shapiro (2008) found that women are motivated by their employment-related needs and one-third (6 out of 18) of the women from Gerosa *et al.* (2021)'s study reported this motivation to join OSS. This is in line with previous research that shows that women are more motivated to use technology for what it enables them to accomplish (Burnett *et al.*, 2010). Reputation was not a top motivation for contributors from any gender in the early 2000s (Lakhani and Wolf, 2003; Ghosh *et al.*, 2002) and is still not. Only 4% of the men from Gerosa *et al.* (2021)'s study joined because of reputation, while no women reported it.

Finally, an interesting case was Ideology, which was top-ranked in general surveys from the 2000s about motivation to join OSS (Lakhani and Wolf, 2003; Hertel *et al.*, 2003; Hars and Ou, 2004; Ghosh *et al.*, 2002). Ideology is usually captured by motivations such as "software should be free for all," "[software should be] free to modify and redistribute," or "OSS should replace proprietary software." (Von Krogh *et al.*, 2012). In the recent survey by Gerosa *et al.* (2021), this motivation has dropped some and was mentioned as a reason to join by only 11% of the women and 11% of the men. When considering only women, we can see that there was also a drop compared to the OSS survey, in which ideology was mentioned by 28% of the women (Robles *et al.*, 2016).

Reciprocity, kinship, and pay are especially relevant motivations for women. On the other hand, enjoyment and fun motivate more men than women. Altruism, learning, and own-use motivate both men and women, career and reputation are low-ranked for both genders, and ideology was relevant in the past but has lost importance for both genders over the years.

3.3 What types of contributions do women make in OSS projects?

The OSS landscape has changed since the early 2000s to include the participation of ever more people and companies. Project-centric roles are becoming more established, and OSS projects increasingly include community-centric roles, which relate to areas beyond programming (Trinkenreich *et al.*, 2020a). Understanding how women contribute to OSS projects can help attract both code developers as well as those interested in being part of OSS in a non-code-centric role.

While the primary type of contribution in OSS projects is related to code development, the roles available in OSS projects go beyond the project-centric roles Trinkenreich *et al.* (2020a). They include many people who work "behind the scenes" to drive and sustain the community (Trinkenreich *et al.*, 2020a). We categorized the activities reported by primary studies according to a framework of OSS roles from a previous study (Trinkenreich *et al.*, 2020a) with two perspectives: coders or non-coders, and project-centric or community-centric (Fig.3.5). Next, we present details about women's contributions as coders and non-coders reported by primary studies.

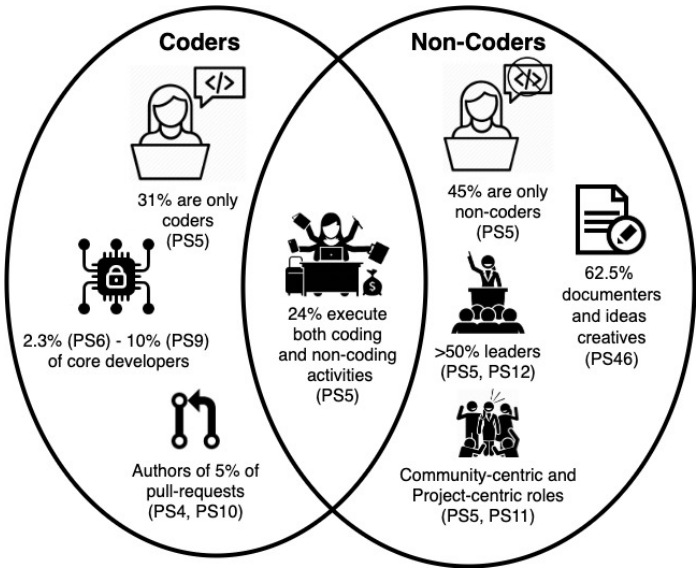

Figure 3.5: Women make OSS contributions both as coders and as non-coders

3.3.1 Coders

Only 31% of the 226 women who answered the OSS survey Robles *et al.* (2016) contribute to OSS projects as code developers, and 24% perform coding in parallel with other roles.

Core Developers. The classic hierarchical model of coders in OSS development communities is described as a core-periphery structure, with a small number of core developers and a large set of peripheral developers (Nakakoji *et al.*, 2002). The core developers are code contributors involved with the OSS project for a relatively long time who make significant contributions to guide the project's development and evolution. Due to their relevant contributions and interactions, core developers often play leadership roles in OSS projects (Ye and Kishida, 2003). Related to these developers, Canedo *et al.* (2020) found women as core developers in only 5.24% of the 711 GitHub analyzed projects. Of all the core developers, only 2.3% were women. From a dataset of 683,865 code review requests

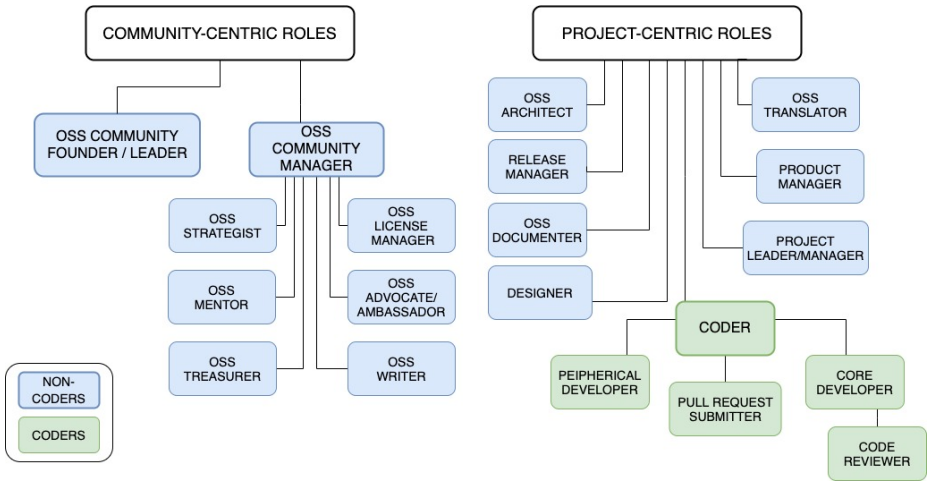

Figure 3.6: Community-centric and project-centric roles reported by primary studies as played by women who contribute to OSS. Roles can be played in parallel. Adapted from (Trinkenreich *et al.*, 2020a)

from ten popular OSS projects, Bosu and Sultana (2019) found that women comprise a maximum of 10% of core developers among all ten projects. Following a classification of commit types from Hattori and Lanza (2008), Canedo *et al.* (2020) concluded that women who are core developers contribute more with corrective and reengineering commits than forward engineering and management commits. Moreover, when describing the commits, women present a more detailed message explaining their contribution changes than men. El Asri and Kerzazi (2019) evaluated the interactions in projects Angular.js, Moby, Rails, Tensorflow, Django, Elasticsearch and found that women core developers are more likely than men core developers to interact with other contributors, evolve similarly to men within the project, and, though underrepresented, contribute to building sustainable social capital for OSS.

Preferred Technologies. From the 711 projects analyzed by Canedo *et al.* (2020), women represented 8.8% of the core developers in projects that are based on the *Scala*

programming language, 8.7% with *CSS*, 6.3% with TypeScript, 5.6% with Swift, only 1% when the project is based on PHP and Shell programming languages. Still, from the same data, even projects written in TypeScript (10.93%) have at least one woman core developer, and 2.17% of projects using PHP have at least one woman core developer. Considering the data from the six projects analyzed by El Asri and Kerzazi (2019) (Angular.js, Moby, Rails, Tensorflow, Django, Elasticsearch), women represented at least 4.8% of core developers for the projects based on Python, at least 4.5% in C++, and at least 4.2% in Java. From the 158,464 of pull-requests for which Terrell *et al.* (2017) could identify gender, women had a greater rate of accepted pull requests in Ruby, Python, and C++. We present the percentages of women as developers for each programming language as reported in the primary studies in Fig 3.7.

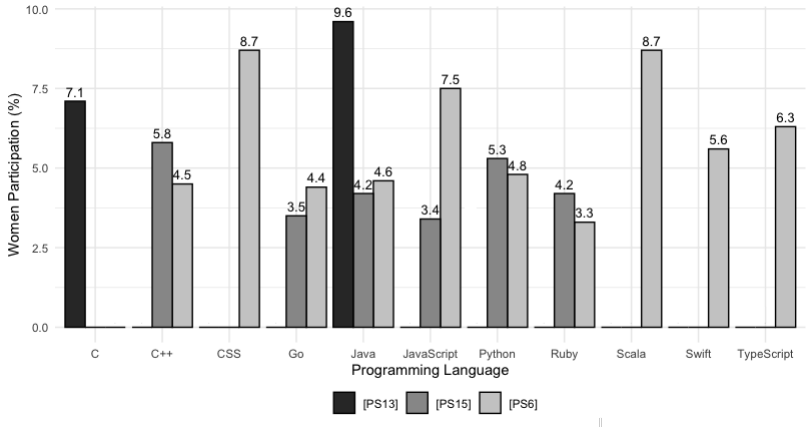

Figure 3.7: Women's participation as core developers in OSS per programming language

Code reviewers. Peer code review is a practice in software engineering. A code developer submits the code produced to another person (peer) to evaluate and find possible errors before merging the code to the project codebase (Bacchelli and Bird, 2013). According to the GitHub dataset of six projects evaluated by Paul *et al.* (2019), women are more likely to write reviews expressing sentiments in the text to another woman than to

a man during code reviews. Huang *et al.* (2020) used medical imaging and eye-tracking to evaluate the visual and cognitive processes and patterns of neural activation followed by reviewers while performing code reviews. Authors found that women spent significantly more time analyzing pull-request messages and author pictures (regardless of their identity) than the code itself when performing code reviews.

3.3.2 Non-Coders

Almost half (45%) of the 226 women who answered the OSS 2013 survey take part in non-coding activities, and 24% perform code-related activities in parallel with other roles (Robles *et al.*, 2016).

OSS Community Manager. After analyzing the career pathways followed by 17 contributors, Trinkenreich *et al.* (2020a) presented a set of community-centric roles, including community founders and managers, strategists, mentors, writers, license managers, treasures, and advocates. The contributors who play these roles are usually "hidden figures," who are not visible when analyzing the data from projects' repositories or coding platform websites. 11 (out of the 12) women interviewed in this study play community-centric roles.

Project Leader/Manager. More than half (51.49%) of the 226 women who answered the OSS 2013 survey participate in community leader, coordinator, or administrator roles, while only 5% of those women coordinate more than three projects (Robles *et al.*, 2016). Women in OpenStack who play leadership roles represent 7% of the project committee members, 8% of the project team leaders, 9% of the project board directors, 7% of the technical committee, 8% of the working group leaders, and 23% of project ambassadors (Izquierdo *et al.*, 2018).

Women are more present in community-centric than in project-centric roles; almost half make non-code contributions. However, they often play both coding and non-coding roles in parallel. Very few projects have at least one woman as a core developer.

3.4 What challenges do women face when contributing to OSS projects?

Previous work investigated challenges faced by OSS contributors who are mentors and newcomers Steinmacher *et al.* (2015b); Balali *et al.* (2018); Mendez *et al.* (2018a,b); Steinmacher *et al.* (2016). Some of them report gender bias as a challenge Balali *et al.* (2018), and others report barriers faced by women Dias Canedo *et al.* (2019). Anecdotes about gender bias appear across the literature (Terrell *et al.*, 2017; Kofink, 2015), and women have been reported to feel that such biases are to blame for their contributions' comparatively low acceptance rate (Canedo *et al.*, 2020).

We aggregated the scientific evidence about challenges and gender bias. Understanding the nuances of women's challenges to contribute to OSS can help communities plan strategies to mitigate these challenges and thereby attract and retain more women.

Women mainly face socio-cultural challenges when contributing to OSS (Lee and Carver, 2019), which can also influence their decision to leave an OSS project Paul *et al.* (2019). From the 37 women who answered Powell *et al.* (2010)'s survey, 50% of them indicated they had witnessed gender-based discrimination within the OSS community either online, in meetings, or in class, and 50% said they had experienced harassment online or offline. Gender-related incidents can be so severe that they motivate women to leave an OSS project (Vasilescu *et al.*, 2015a). Indeed, in another survey (Lee and Carver, 2019) women reported that they drop out when the OSS project does not care about diversity.

Leaving an OSS project is a decision that impacts more women than men—according to Qiu *et al.* (2019c), women are more likely to disengage from GitHub by 27%. Understanding the reasons behind the decisions to step away from a project can help create strategies to increase retention in OSS. Kuechler *et al.* (2012) suggest that women drop out because the OSS project is not aligned with their motivations or due to unappealing and hostile social dynamics.

We summarize the challenges found in the literature, which were all socio-cultural, in Fig. 3.8, and mark with an asterisk (*) the ones reported as a challenge that ultimately can cause women to leave OSS. Next, we present and explain each of them.

Figure 3.8: Challenges faced by women when contributing to OSS

Lack of Peer Parity. Most women (72%) feel outnumbered and 24% feel alienated (Powell *et al.*, 2010). Women reported feeling more comfortable and accepted by their same-gender counterparts (Balali *et al.*, 2018) and feel frustrated when there is no peer parity (Vasilescu *et al.*, 2015a). This problem worsens in medium-size projects in which all contributors are men, as they may form a clique that a woman contributor could have difficulty breaking into (Qiu *et al.*, 2019a). *"It is not so common to find many girls in technical teams"* (Canedo *et al.*, 2020), and in the face of this lack of parity women reported feeling

invisible in larger majority-men groups (Calvo, 2021).

Non-inclusive communication. Discriminatory expletives, swear words, and negative critiques often used in code reviews and mailing lists may be insulting to women. The negative workplace experience of encountering words that are demeaning to them in mailing lists can cause women to leave OSS projects (Paul *et al.*, 2019). Awkward communication styles (Balali *et al.*, 2018), acrimonious talk about which code piece should be incorporated (Nafus, 2012), and terms usually associated with men (e.g., "guys") can demotivate women (Qiu *et al.*, 2019a).

Toxic Culture. Incidents of symbolic violence and harassment against women can hinder their access to the community, as when men decide to 'Hire that one because she is hot' (Calvo, 2021). Geek-saturated communities like Slashdot are often unwelcoming and hostile environments (Parker, 2000). Additionally, women are sexualized in OSS (Singh, 2019b), facing judgment, abuse, hostility, and discrimination (Singh, 2019b). While hurtful and offensive talk is openly addressed to them, women are obliged to remind men not to "stare and point" at them (Nafus, 2012). From the 13 women who answered a questionnaire sent to 15 OSS projects, 38% had suffered some incidents of sexism, including sexist statements or assumptions, being ignored, insinuations that they had it easy because they were women, and being simultaneously held to higher standards than men and underestimated (Lee and Carver, 2019). Moreover, the authors found that women had trouble being taken seriously and needed to prove themselves (prove-it-again (Imtiaz *et al.*, 2019)). According to Kuechler *et al.* (2012), public flaming and aggression can be enough to distort women's participation, as they may already be hesitant about how they will be received.

Problems of Impostor Syndrome. Women tend to be risk-averse (Dohmen *et al.*, 2011) and have low computer self-efficacy (Burnett *et al.*, 2010, 2011; Cazan *et al.*, 2016;

Hartzel, 2003; Huffman *et al.*, 2013; Singh *et al.*, 2013), which can affect their behavior with technology. Compounded with lack of safety due to a toxic culture, women can become less confident in their ability to complete tasks and blame themselves if there is a problem (Mendez *et al.*, 2018a,b; Padala *et al.*, 2020). Women find it challenging to directly translate competence to confidence without social attraction (being liked by the other community members in terms of having a rapid increase of followers). Consequently, initiating a pull request to a new repository can be problematic due to women's competence-confidence gap Wang *et al.* (2018). Even understanding that confidence is an essential factor when entering OSS Powell *et al.* (2010), women face a lack of self-efficacy (Lee and Carver, 2019; Balali *et al.*, 2018). *"Despite having proved [their] competency in certain areas of the code/project, [their] opinion is rarely or never asked for"* (quotation from Vasilescu *et al.* (2015a)). Still, Imtiaz *et al.* (2019) found that women tend to be more restrained than men in general. Despite being knowledgeable and professionally well-settled, women may be more reluctant to publicly display their work (Vasilescu *et al.*, 2015a).

Community Reception Issues. Women reported feeling restrained when communities ignore them when they lack the skills to provide contributions on their first day (Moon, 2013). When trying to find a mentor, upon discovering their mentee's gender, men mentors can treat the relationship as a dating opportunity (Nafus, 2012). This makes finding a mentor an arduous task, which includes attracting attention and breaking into a close-knit OSS community (Lee and Carver, 2019). Many women use fake GitHub accounts and hide their gender, *"so that people would assume [they] were male"* (Vasilescu *et al.*, 2015a).

Stereotyping. Pre-existing stereotypes (Singh, 2019b; Powell *et al.*, 2010), gender roles, and "macho" attitudes can cause gender inequalities in OSS communities (Calvo, 2021). Women are boxed into specializations despite their manifest protest against it, as

the legal case against the front-end/back-end distinction has shown (Vedres and Vasarhelyi, 2019). Additionally, men often treat women as if they were their mother, asking for advice about how to dress and behave and then refusing to enter into a technical dialogue thereafter (Nafus, 2012).

Work-life Balance Issues. Women that participated in Lee and Carver (2019)'s study reported a lack of time and family responsibilities. Only women from this study reported family responsibilities as a challenge.

Gender-Biased Peer-Review. Even in a population of core developers, one-third of the 36 women who answered Canedo *et al.* (2020)'s survey reported they believe that reviewers had not accepted at least one of their contributions due to gender bias. Moreover, 11.4% of the women participants perceive gender bias while someone review their contributions. Although women can have a merge acceptance rate nearly equivalent to or a little higher than men's (Terrell *et al.*, 2017; Powell *et al.*, 2010), according to Terrell *et al.* (2017) there is a bias against women's contributions when their gender is known. The authors found that women have a 12% lower acceptance rate when they explicitly identify themselves as women, compared to 3.8% for men who disclose their gender. However, authors found that women tend to have their pull requests accepted at a slightly higher rate (78.7%) than men (74.6%) when not identifying the gender, regardless of experience level. While less experienced developers making their initial pull requests do get rejected more often, women generally still maintain a higher rate of acceptance throughout. Bosu and Sultana (2019)'s study corroborated the bias that women face by using three additional metrics (first feedback interval, review interval, and code churn per comment). The study analyzed ten projects and found explicit biases against women in three of the analyzed projects (Android, Chromium OS, and LibreOffice). Women had lower code acceptance

rates than men and had to wait longer to receive initial feedback for their code changes and to complete code reviews. The code submitted by women also had lower churn per comment in both Android and Chromium OS. The study showed that Android and LibreOffice stood out in having prominent gender biases, where women had 10% lower acceptance rates than men, and review intervals that last three times longer than men. On the other hand, three other projects indicated biases favoring women (oVirt, Qt, and Typo3).

Women face many socio-cultural challenges that include: missing being around other women (lack of peer parity), offensive language in mailing lists (non-inclusive communication), suffering symbolic violence and harassment (toxic culture), avoiding initiating a pull request due to feeling unsafe and lacking confidence (problems of impostor syndrome), facing challenges asking in finding mentors (community reception issues), being boxed into specific roles (stereotyping), splitting their time between work and family (work-life balance issues), and biases or lower acceptance rates of contributions when they explicitly identify themselves as women (gender-biased peer review).

3.5 What strategies were proposed to mitigate the challenges and support women's participation in OSS projects?

Strategy recommendations to improve diversity are scattered and rarely widely adopted. OSS communities need a concise view of the different types of actions to select the ones that are viable and appropriate for their needs and for the challenges they face. Strategies include actionable mechanisms that OSS communities can adopt and combine to create a more inclusive environment for women in OSS.

We summarize the proposed strategies in Fig. 3.9. Next, we present and explain each of

them. All strategies were mentioned as a way to mitigate at least one challenge presented in Sect. 3.4. However, there was no strategy reported to mitigate either the challenge Work-Life Balance Issues or Gender-biased Peer Review.

Promote awareness of the presence of peers. Approximately half (54%) of the women respondents of Powell *et al.* (2010)'s study said they would be more inclined to participate in OSS if there were more women involved. Promoting awareness about the rate of women can help to attract more women, minimizing the feeling of being the only one in the room and increasing a sense of safety (Powell *et al.*, 2010). This awareness should include a measurement of women's participation and the type of contributions. According to Calvo (2021)'s study, the communities can generate parallel spaces in which the proportion of women is above 50% so as to create more diverse environments under the values of mediation and care.

Promote women-specific groups and events. The community managers interviewed by Calvo (2021) mentioned that they promote schoolgirls' events to inspire vocations and empower girls who may opt for an OSS career. This strategy should promote activities exclusively for women in those spaces, highlighting their presence, as women tend to be invisible in larger groups in which men comprise the majority (Calvo, 2021). For those women already interested in OSS, promoting women-only groups, spaces, and events (Canedo *et al.*, 2020; Singh, 2019b; Calvo, 2021; Singh, 2019b; Powell *et al.*, 2010; Singh and Brandon, 2019) fosters discussions, supports networking, and fuels empowerment (Singh, 2019a). Moreover, it provides a safe space for expressing feelings and opinions (Calvo, 2021) and revealing their identities (Singh, 2019a). Although effective, Singh (2019a) found that only 3% of the 350 projects they analyzed have women-specific spaces—including websites, IRC Channels, dedicated blogs, collection/list of resources, dedicated

Facebook pages, and/or local meet-ups.

Promote inclusive language. Avoid gender pronouns that assume that people are [all] one gender or one demographic (Wurzelová *et al.*, 2019). For example, using 'guys' is common, and can give an impression that contributors are men (Canedo *et al.*, 2020).

De-stereotype the OSS contributor. The women interviewed in Singh (2019b)'s study recommended to "leave the stereotypes out the door." The frustration caused by stereotypes was expressed by one of the women surveyed by Canedo et al. (Canedo *et al.*, 2020): "Stop treating women developers as 'women developers' and start treating them as developers." Powell *et al.* (2010) suggests showing less discrimination and more inclusion to tone down the male-dominated atmosphere and promote participation. Calvo (2021) and Vedres and Vasarhelyi (2019) suggest avoiding the *feminization* of specific assignments, like those relating to community building tasks; OSS communities should reclassify types of work that have been packaged in masculine-feminine stereotyped specialties.

Encourage and welcome women. Singh (2019b)'s findings show that being less judgmental and appreciating diverse teams is essential to supporting and encouraging women. Indeed, as Beach (2014) discusses, people need to feel supported, accepted, and encouraged. This encouragement may even come from other women (Prana *et al.*, 2021). Powell *et al.* (2010) suggested starting by encouraging small steps as an incentive to women to submit bug reports and share their input, which was echoed by two participants of Canedo *et al.* (2020)'s study, "the solution is to build confidence" and "not to fear when contributing." This would increase their self-confidence Parker (2000). Encouragement is also the goal of some initiatives presented by Parker's Parker (2000) FLOSSpols [3] , which offers

[3] www.flosspols.org

recommendations on how to solve the gender gap; these initiatives include WOWEM, a gender equity and OSS research and education project; and LinuxChix, a community for supporting women in Linux. There is no value in encouraging women to be there if the environment is hostile. To welcome women, one of the community managers who participated in Barcomb *et al.* (2020)'s study recommends making the community friendlier in general.

Promote women to leadership roles (empowerment). A way to empower women is to have them in senior roles (Canedo *et al.*, 2020), in project governance (Qiu *et al.*, 2010), and where appropriate (Singh, 2019b), as mentioned by one woman participant of Prana *et al.* (2021)'s study: "More women reviewers. More women are acting directly on the governance of large OSS projects". Some community managers described how their communities created decision-making positions and ensured that they held women-led public activities (Calvo, 2021). Catolino *et al.* (2019) suggested that a way to avoid the proliferation of community smells is to involve women in positions where they can mediate discussions and improve the communication of sub-communities. According to Singh (2019b), promoting women to positions of authority shows the project respects their contributions. De-bias tools. Most (73%) of the barriers that affect software professionals have some form of gender bias (Mendez *et al.*, 2018a). Indeed, bias in tools and infrastructure can hinder women who are newcomers from joining OSS (Mendez *et al.*, 2018a). One way to de-bias infrastructure and tools is by applying the GenderMag technique. GenderMag uses personas and a specialized Cognitive Walkthrough (CW) to systematically evaluate software and make them more inclusive of the women's cognitive styles (Mendez *et al.*, 2018a,b; Padala *et al.*, 2020). This technique's precision was proved by a lab study that showed that the GenderMag technique helped to identify 81% of the issues (Burnett

et al., 2016).

Recognize women's achievement (visibility). When the community recognizes women's achievements, it provides the social attraction that women seek to overcome their competence-confidence gap (Wang *et al.*, 2018). Communities can show recognition by increasing the visibility of women (Prana *et al.*, 2021), listing them as great contributors whenever they deserve it (Singh, 2019a), and *publicly* celebrating their achievements in blogs, project homepages, and social media (Imtiaz *et al.*, 2019; Prana *et al.*, 2021). Another way to increase visibility is to organize events where speakers are women (Calvo, 2021; Singh and Brandon, 2019). These simple actions inspire more women to participate Prana *et al.* (2021); Calvo (2021).

Prepare Mentors to Guide Women. As mentioned before, OSS projects are usually men-dominated environments, which may scare women away (Nafus, 2012; Turkle, 2005b). Mentorship can help women newcomers find the assistance and support they need Kuechler *et al.* (2012); Calvo (2021); Barcomb *et al.* (2020). One way to do so is to ensure that women are in positions to mentor other women (Singh, 2019b,a). Singh (2019b) underscores that when mentoring women it is necessary to guide them on different aspects. While men need to change their behavior and projects need to implement systemic changes, Singh (2019b) posits that women also need to be trained to ignore disruptions and not be easily bothered by criticism or insults; the mentor needs to be extra supportive, friendly, respectful, and encouraging (Singh, 2019b).

Create and enforce a Code of Conduct. Developing a code of conduct for the community (Imtiaz *et al.*, 2019; Singh, 2019b; Prana *et al.*, 2021; Calvo, 2021; Wurzelová *et al.*, 2019; Tourani *et al.*, 2017; Fossatti, 2020; Singh and Brandon, 2019) helps to mitigate

tightrope effects [4] in articulating acceptable behaviors for all members (Singh, 2019b). The code of conduct comprises the collective norms of a community, as mantras to shape the culture of collaboration (Fossatti, 2020), including the community's expectations and values to create a friendly and inclusive community (Singh, 2019b; Tourani *et al.*, 2017). While having a code of conduct will not prevent sexism, it indicates to any men who have firmly held anti-female behaviors that such actions will not be tolerated in the project (Lee and Carver, 2019). However, according to Robson (2018)'s study, just creating a code of conduct will not increase women's participation. The author showed that projects that introduced a code of conduct in their history saw women's participation increase almost at the same rate as projects without a code of conduct. Projects increased from 2.37% to 3.81% after a code's introduction. Projects without codes of conduct, comparing gender diversity within similar periods yields, had an increase from 4.10% to 5.53%. The average increases were 1.44%, and 1.43%, respectively, which means creating the code of conduct did not help increase women's participation. Robson (2018) posits that the code of conduct needs to be enforced among the project members. Indeed, it is necessary to have mechanisms in place to implement the code and show that violations have consequences (Singh, 2019b; Tourani *et al.*, 2017).

[4]The term tightrope is usually associated with the circus, where a performer carefully balances while walking across a narrow stretched rope suspended in the air. As an analogy, women behave in a restrained manner to avoid backlash. The term can refer to the narrow band of socially acceptable behavior for women (Imtiaz *et al.*, 2019) by assisting communities

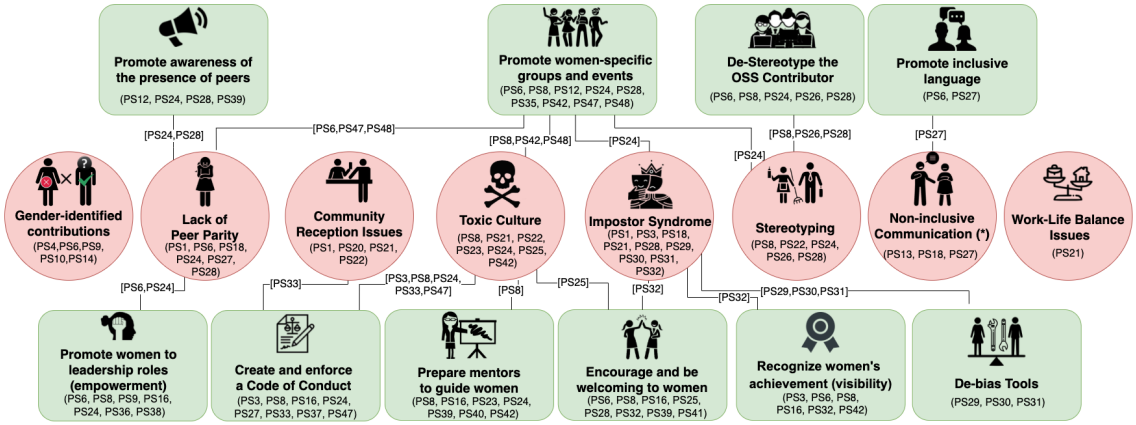

Figure 3.9: The strategies that were mentioned by primary studies to mitigate challenges faced by women in OSS (circles represent challenges and rectangles represent strategies)

The strategies suggested by literature for OSS communities to employ and increase women's participation include providing awareness and numbers about women contributors, inclusive language, women-specific groups and events, de-stereotyping the OSS contributor, encouraging and welcoming women, having women in leadership, de-biasing tools, recognizing women's achievements, preparing mentors to guide women, and creating and enforcing a code of conduct.

3.6 Challenges in improving diversity

3.6.1 The leaky OSS pipeline

Section 3.1 reports that there is a large gender disparity in OSS contributors (Kofink (2015) found less than 5% of pull-request authors were women). However, the gender disparity is less pronounced in the initial stages (e.g., for students of Google Summer of Code). There is attrition of women contributors as they move through the different stages of the "joining script" (Von Krogh *et al.*, 2003)—where people start outside the project as

readers and passive users; then move to the project periphery as bug fixers, bug reporters, peripheral developers, and active developers (developers without commit rights); and finally, enter the project as core members or project leaders (Nakakoji *et al.*, 2002)—the leaky pipe phenomenon. This attrition can be a consequence of the several socio-cultural challenges faced by women during the process. As we presented in Section 3.4, women face gender bias in communication and community acceptance, and lower contribution acceptance rates when they explicitly identify themselves as women. While mentorship events enhance (women) participants' sense of competence and increase the chances of future contributions' values Silva *et al.* (2020a), these programs alone are insufficient, as women do not stay long enough to become project leaders. The majority of the challenges that women face or the reasons women leave OSS (see Section 3.4) are socio-cultural in nature and unrelated to technical skills. Therefore, strategies that help create an inclusive environment geared towards retaining and mentoring women are needed to fix the leaky pipeline.

3.6.2 Problems of impostor syndrome caused by systemic oppression

Impostor syndrome is a psychological concept about a pattern of behavior wherein people (even those with adequate external evidence of success) doubt their abilities and experience a persistent fear of being exposed as a fraud (Mullangi and Jagsi, 2019). Impostor syndrome entered the mainstream as a buzzword. As we presented in Section 3.4, the literature shows that when women join OSS, the rate of acceptance of their contributions is similar to—if not higher than—men's. The systemic oppression of a Toxic Culture has a cumulative effect and disproportionately affects women and other minority groups, who must constantly prove themselves Imtiaz *et al.* (2019). They can develop

unrealistic expectations for their performance and overwork themselves to the point of physical and mental strain Ramos and Wright-Mair (2021). However, the need to prove-it-again Imtiaz *et al.* (2019) and the confidence gap happens due to systemic oppression, not due to an individual's inadequacy. Women are not the problem, rather the system is Ramos and Wright-Mair (2021). Combining actions that include training for allies with both reactive and proactive mechanisms to enforce the code of conduct among the project members Robson (2018) can help mitigate the toxic culture, as we present in Section 8.2.

3.6.3 Impact of gender stereotyping

Research shows conflicting findings about the effects of diversity on team performance. Although low diversity can enhance mutual trust and effectiveness, demographic similarity may also lead to stereotyping, cliquishness, and conflict (de Gilder and Wilke, 1994; Molleman and Slomp, 2006). Regardless of the cause, according to a survey of 5,500 GitHub users (Zlotnick, 2017a), women more often than men encounter language or content that makes them feel stereotyped. Stereotypes manifest common expectations about members of certain social groups. Both descriptive (how women are) and prescriptive (how women should be) gender stereotypes and the expectations they produce can compromise a woman's career progress (Heilman, 2001, 2012). Even before starting a career, stereotype threats represent one of the significant barriers to underrepresented groups engaging in Computer Science education. Implicit stereotypes about gender and STEM have profound effects on girl's or women's interest, confidence, and persistence in STEM education and careers (Dasgupta and Stout, 2014; Dasgupta, 2011).

Fear of gender stereotyping can lead women to hide their gender (Vasilescu *et al.*, 2015a) and create pseudonyms to avoid judgment (Lee and Carver, 2019). This behavior

was also observed by Ford *et al.* (2019) in online communities, where participants use a "gender neutral alias for websites like technical communities, because [they] get better help when asking questions or answering them." One of the strategies presented in Section 3.5 is De-stereotyping the OSS contributor, which differs from using a neutral username to hide gender. In fact, Canedo *et al.* (2020)'s study showed that users who do not reveal their gender suffer an even more severe disadvantage in survival probability. Although it prevents discrimination by categorical gender, avoiding gender identity can lead to a lack of trust and exclusion from projects and ultimately cause a higher exit rate for such users. Even when stereotyping is minimal, it can still make a difference. People's attitudes, beliefs, and behavior are often shaped by factors that lie outside their awareness (Banaji and Dasgupta, 1998; Greenwald and Banaji, 1995). The stereotype threat mechanism has been proven as an effect that harms girls' abilities to solve mathematics tests. When saying the problem is complex, and boys can solve it faster, it induces an antipathetic effect and emotional and cognitive overload, undermining girls' outcomes as opposed to relaxed males Spencer *et al.* (1999); Tomasetto *et al.* (2011).

Previous research showed that women are more present in community-centric roles playing non-code activities and represent less than 5% of coders in core positions of the OSS projects 3.3.2. Although we bring this information and show the research about women's rates in OSS projects, the current reality should not perpetuate a descriptive stereotype about women being less suitable or capable of playing coding activities.

Considering that even minimal social cues may activate negative stereotypes early in informational processing (Wu *et al.*, 2020), De-stereotyping the OSS contributor is crucial for women to start seeing themselves playing the role of developer and not just men. This strategy is aligned with the suggestion of women interviewed in Blincoe *et al.*

(2019)'s study, who considered that changing the typical image of software engineers as IT geek men is a way to reduce the gender gap.

3.7 Concluding remarks

Our literature mapping has shown severe gender disparity in OSS, with women's representation at about 10%. Further, the women who participate are generally volunteers who can devote less than one workday a week to OSS. Although present in community-centric roles, women are less likely to be authors of pull requests and core developers, with many making non-code contributions. Gender biases exist in OSS in several places. For example, when submitting a pull request, women generally have high rates of merge acceptance, but lower rates when they explicitly identify themselves as women. Women also face social challenges such as a lack of peer parity, non-inclusive communication, a toxic culture, impostor syndrome, community reception issues, stereotyping, work-life balance issues, and bias against gender-biased peer review. OSS communities that seek to increase women's participation can mitigate these challenges by providing awareness about the presence of other women, promoting inclusive language, organizing women-specific groups and events, de-stereotyping the OSS contributor, encouraging and welcoming women, placing women in leadership, adopting de-biasing tools, recognizing women's achievement, preparing mentors to guide women, and creating and enforcing a code of conduct. In the following chapters, we report primary studies executed with OSS contributors, starting with the career pathways and different types of contributions.

Chapter 4

CAREER PATHWAYS AND TYPES OF OSS CONTRIBUTIONS

OSS projects need people in both technical and non-technical roles and activities to keep the project sustainable and evolving (Steinmacher *et al.*, 2017; Robles *et al.*, 2019). Women are more present in community-centric than in project-centric roles, and almost half make non-code contributions (Robles *et al.*, 2016; Trinkenreich *et al.*, 2021c). However, OSS communities have been studied as technical communities, where stakeholders join and evolve in their careers based on their code contributions to the project. The OSS landscape is changing, with more people and companies getting involved. In this context, we investigated the roles and activities that are part of the current OSS landscape and the different career pathways in OSS.

Only 4.07% of the 226 surveyed women from an OSS 2013 study joined to increase their job opportunities. After becoming contributors, this motivation increased almost six times (to 25.79%) (Robles *et al.*, 2016). This can represent the "shifting belief" that women have in OSS toward building a career, which increases only after overcoming the barriers to joining and becoming contributors. An awareness of the different roles and career pathways that exist in OSS can attract women with diverse backgrounds and expertise to OSS by showing them the multitude of trajectories to success.

This study was published at the 23rd ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW) in 2020 and received the Honored Mention

award (Trinkenreich *et al.*, 2020b).

4.1 Method

We interviewed 17 people who are well-known in OSS (12 of them identified their gender as women) that were invited speakers at OSCON 2019 (Open Source Software Conference).

We asked about the story of their career, how they joined, which roles and activities they perform, and how they arrived at their current position in OSS. Through qualitative analysis of the career story told by our interviewees, we identified roles that do not use nor contribute to OSS (Not-related to OSS), roles in which OSS is being used or consumed (OSS Consumer), and roles in which OSS receives a contribution (OSS Contributor).

The research design for this study is presented in Figure 4.1.

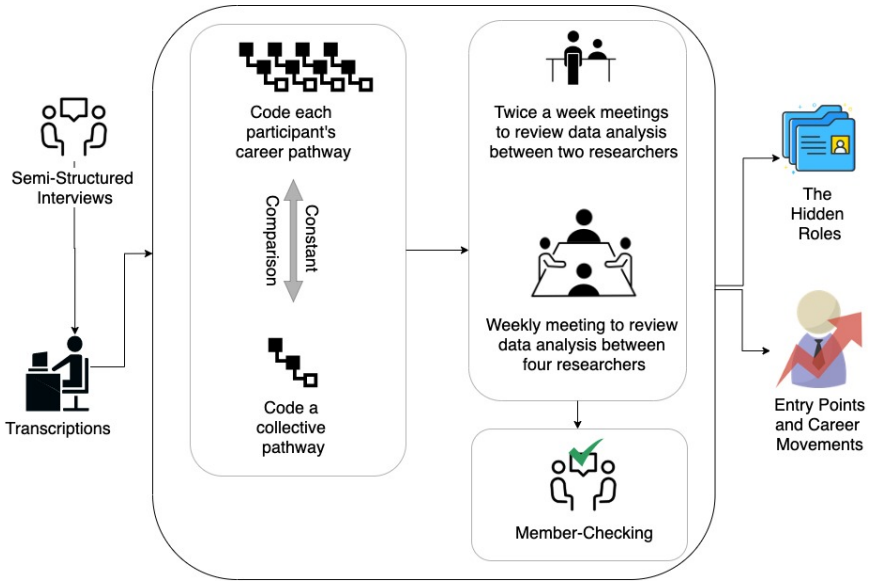

Figure 4.1: Research design for the study about different pathways and types of contributions in OSS

4.2 Results

Figure 4.2 shows that the OSS roles reported by our participants are diverse and are taken up during different stages of contributors' careers. In this figure, we summarize the roles and transitions between roles in the career pathways reported by participants. We used an * to represent people who simultaneously work in more than one role.

We found there are different ways to contribute to OSS, which can take different forms, as aptly mentioned by P3: *"all of the team members there brought this insatiable curiosity about what the others around the table had to contribute ... contribution can take three forms: time, talent, and treasure ... ".*

We then separated the OSS Contributor roles into two groups, according to the activities reported by our participants: Community-centric roles and Project-centric roles.

In addition to the common project-centric—mostly code-related—roles (*e.g.*, coder, system admin, and project manager), our analysis identified the emergence of a set of community-centric—including non-code-related—roles and activities (*e.g.*, advocates and mentors). Although some of these roles are common in the software industry, they were not common (and many are still not formally recognized) in OSS until recently. In this section, we will present these roles, as reported by our participants, starting with the community-centric ones, and depict their importance to the current landscape of OSS.

Community-centric roles This category includes roles related to: (a) the creation of the community (i.e., OSS Community Founders) and (b) the management of the community (i.e., OSS Community Managers). These roles, although known and important for the projects' sustainability and community evolution, are often not formally recognized

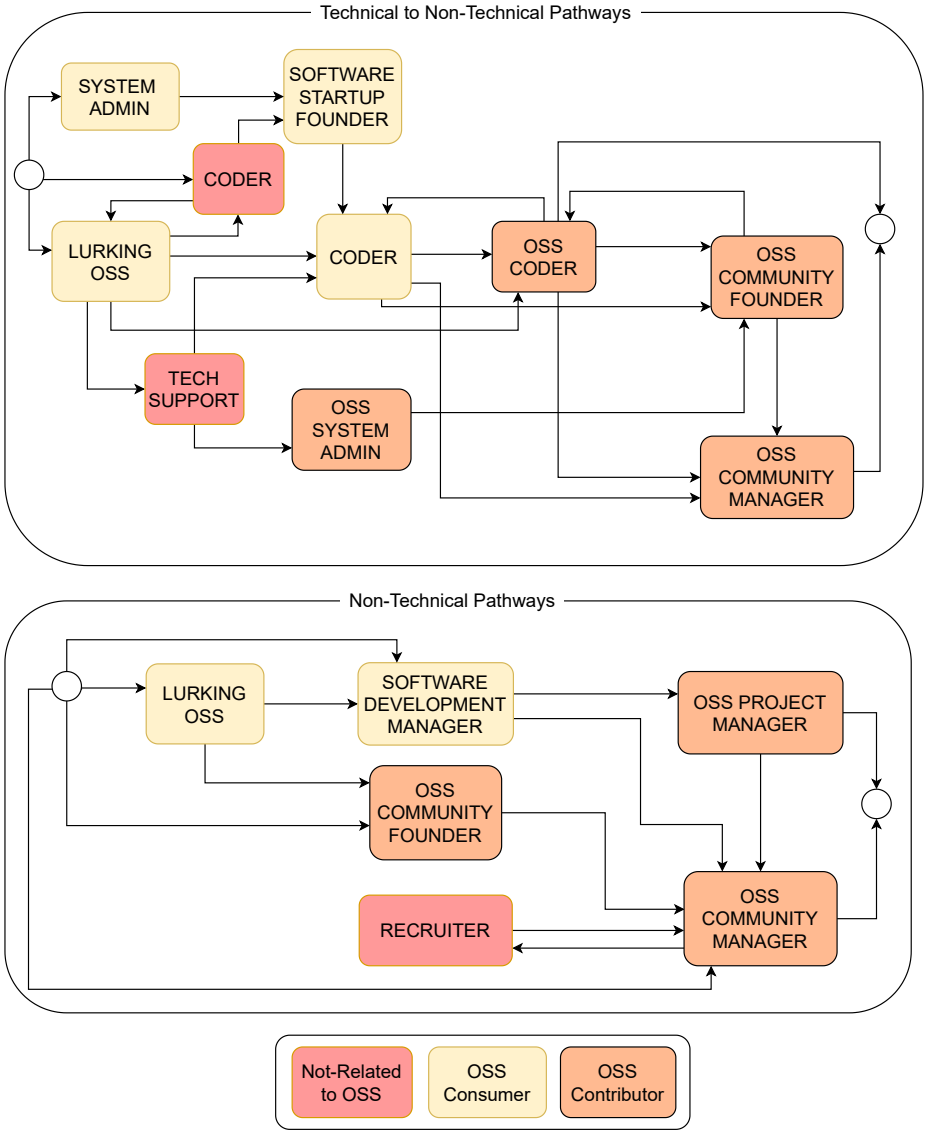

Figure 4.2: Overview of career pathways reported by interviewees. The roles are classified according to their relationship to OSS projects: OSS contributor (pink), OSS Consumer (yellow), and Not-Related to OSS (orange). The labels presented in the arrows represent the participants who took that path (those decorated with '*' parallel roles).

in OSS communities. We present a summary of the roles we identified in Figure 4.3 and discuss them as follows.

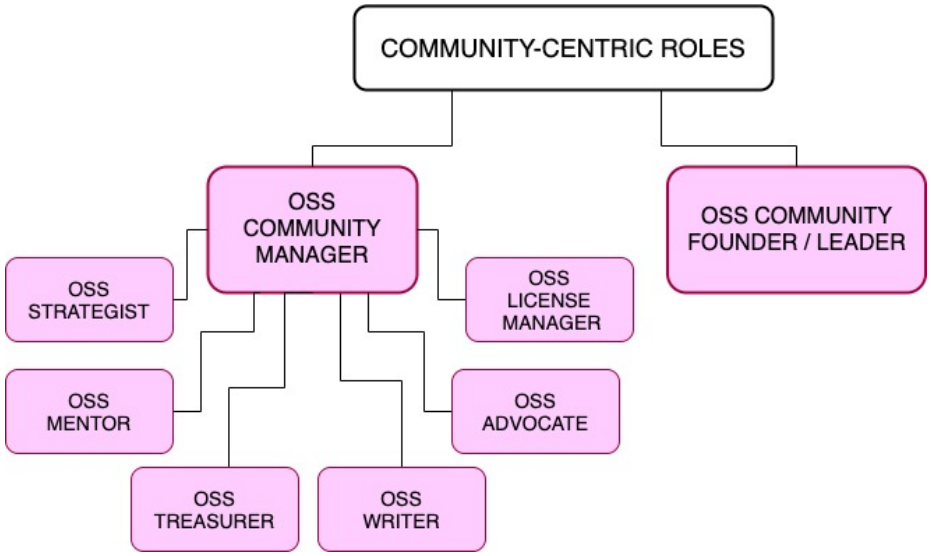

Figure 4.3: The Community-Centric Roles: OSS Community Founder and OSS Community Manager (and the roles that are part of it)

OSS Community Founder: The creation of a new product or project is what truly founds OSS communities, and this role can bring relevant experience to a career in OSS. This role is related to entrepreneurship, but in the OSS landscape, which can benefit both the founder and the community around the product. P4 for example mentioned that *"the bulk of my experience from working with open source communities comes from launching open source community."* We also noticed that "OSS entrepreneurship" was important for developers aspiring for professional development—according to P6 it was *"a way of advancing my own career."* He mentioned that it was important to add this OSS experience to his portfolio. Some interviewees mentioned that founding a community served to showcase their careers (P3, P9). P3 created an OSS content management system and was responsible for providing support services for their solution and for other similar ones. P9

took advantage of the knowledge acquired during his Ph.D. to create an OSS startup that provides quality assessment services for OSS projects.

OSS Community Manager: Simply launching a community or project is not enough to create a thriving community. Most community founders, when they created their projects, did not have the suitable experience to manage the "business" and nurture it toward becoming an active and evolving entity (Li *et al.*, 2019). In this sense, community managers play an important role in sustaining and helping a community mature. The community manager category groups different roles focused on a number of activities that are required to sustain and build an OSS community: advocate, mentor, treasurer, strategist, writer, and license manager.

OSS Advocate: Following the idea behind the sentiment *"why would we not do this together because we are better together"* (P1), this role embraces evangelism, developing plans for *"bringing new people in"* (P10), increasing contributions to the OSS project, and making the community inclusive, welcoming, and safe. It involves activities about *"large mass engagement" (P16)*, how to *"connect and inspire people"* (P1 and P13), find and recognize talents, and build a trustful and friendly connection between the industry and the community. Activities in this role also include encouraging the community to learn the coding skills currently required by the market. One of the practical ways to operationalize advocacy is through events that bring people together. Since we interviewed people who were giving talks at an annual OSS conference, most of them mentioned this as a way to contribute to OSS. However, besides giving talks, some interviewees mentioned that they organized events by selecting *"topics (projects, changes, histories, industry evolution), creating meetups to institute some open source flares in there" (P1)*, planning learning groups *"because we're better together" (P1)*, and producing podcasts when there is a new software

release to *"get people on the show to talk about their specialties" (P15).*

OSS Mentor: In OSS, like any project, newcomers need to be trained in different aspects of the project (Balali *et al.*, 2018), such as becoming acquainted to the project architecture, implementation, and feature details; development guidelines; and organizational rules (Canfora *et al.*, 2012). Although an important activity (Fagerholm *et al.*, 2014), it is often not recognized as a formal role in OSS, and those who perform the activities related to it are not formally trained, recognized, or assigned to play that role (Balali *et al.*, 2018). The perception of our interviewees matched that of prior evidence. Participants mentioned that mentoring is related to *"teaching how to forage" (P10)* and is usually a *"one-on-one engagement" (P10)* toward understanding why the person wants to contribute, present the tasks available, and give task recommendations based on newcomer's motivations.

OSS Strategist: *"When a project starts to grow (or evolves) [it] requires governance"* (P7), and it is critical to understand and control the quality of involved processes. This role is usually related to innersourcing initiatives, which is an emerging topic with growing interest from commercial projects that aim to replicate the success of popular OSS projects internally (Stol and Fitzgerald, 2014). The OSS strategist is responsible for fostering the adoption (of OSS technology) or for improving its processes (to match that of OSS) and improve transparency in organizations or communities, making *"strategic decisions for business" (P8)* and *"moving the company in [the] OSS direction" (P12).* This role is new in the software industry, and adds to the OSS landscape by fostering the involvement of the software industry in OSS, and defining conditions for releasing (in-house) projects under OSS licenses. This is an important role in the current age of OSS, in which companies are central stakeholders.

OSS Treasurer: Treasuring is an important financial activity performed by someone

who leads strategical budgetary decisions for the OSS project or foundation, and who creates a *"sustainable model" (P3)* for the community to support the projects. "*Treasure is literally money, literally money…It's not about technical direction at all, it's literally about making sure that this ecosystem can continue forward*" (P3). The increasing involvement of companies in OSS projects and the involvement of paid developers in the development process make this role important to communities. Bringing in and managing donations and income cannot be done in an *ad-hoc* manner, so having someone trained and focused on the activities is necessary.

OSS Writer: Well-written documentation is not only important for the everyday work of the project, but also can help onboard newcomers, create inclusive communities, and represents a meaningful way to contribute back to OSS. Nearly 25% of the OSS community are not highly proficient at reading or writing in English and need careful, clear, and accessible language in documentation Zlotnick (2017b). Written contributions can be technical documentation about the OSS product, like a software installation guide. Written texts can also encompass educational materials, books, presentations, and publications about OSS. Although not related to coding activities, the outcome of this activity can be visible in project repositories—*"I made contributions to Ubuntu documentation, I got involved with the docs team"*(P10)— and in publications— *"I wrote my book, which was great. And I'm glad that happened"*(P8), " *I've written a few books, the most recent and the most complete one is a case study book, it has some theoretical stuff at the beginning. And at the end, it has practical advice about how you can do it in your company*" (P12). The increasing importance of this role is evident; Google started a program called Google Season of Docs [1] , which focuses on enrolling people who are technical writers to OSS projects to improve

[1]https://developers.google.com/season-of-docs

project documentation. Unfortunately, this role is not well recognized and writing-based documentation activities are performed by people playing other roles.

OSS License Manager: Licenses are an important type of documentation for both users and contributors (Zlotnick, 2017b), since they are the legal means used to regulate how software can be copied, changed, or redistributed. The central importance of licenses, and the specific knowledge needed to make the right decision, justify assigning someone as responsible for licensing, including overseeing the compatibility and compliance of software licenses. Currently, this role is under-recognized and licenses are applied by the maintainers, which may lead to inconsistencies and incompatibility (Wu *et al.*, 2015, 2017; Vendome *et al.*, 2015). The importance of this role is clear based on how long P12 spent on this activity: *"[I] was on the board of the OSI for 10 years, helping to run the licensing question."* P14 started working for an OSS foundation because the board knew she *"could handle certain issues, licensing and compliance because of the [projectname]"*, an experience she had accrued before as a Community Founder.

Project-Centric Roles **OSS Coders:** This role is well-known and is the driving force in OSS product development. Since the beginning of OSS, the coding role has been considered the heart of the model. This role includes activities related to developing new code, maintaining existing code, and writing tests. Our data shows two types: contributors involved with core activities who regularly contributed to the project, and those who were occasionally involved with the project (casual contributors). In the former case, these contributors made it their career and reached "success" as coders—becoming long-term contributors or maintainers. In the latter case, the casual (Pinto *et al.*, 2016) contributors eventually offered new patches or suggested fixes or new features. Often times, these con-

tributions were a side product of another activity (*"hey, here's some code that I wrote"* (P7)) or a reason to stay active in the community (e.g., *"producing [a] certain support product"* (P9)).

OSS Project Managers: People in this role perform two broad types of activities. First, management-centric activities, such as being *"responsible for ... effectively project managing, [in] budget, on time and on schedule, [with] appropriate features and keeping the quality levels high" (P14)* with the goal on *"producing open source projects" (P17).* Second, product-centric activities, which include managing releases and project deliverables as per requirements and schedule constraints, and/or being responsible for the product architecture. The architecture piece is "*a technical and complex activity that represents a career advance, as the architect makes core and strategic decisions about the product and reports to the organization's CTO*" (P14). When involved with large technical changes, an architect can also be a negotiator, having to guide, detect, and harmonize issues between a community of project stakeholders and avoid negative influence on system development (Tamburri *et al.*, 2016). While project management is less explored in OSS literature, it was central to Ye and Kishida's model (Ye and Kishida, 2003). However, they associated this role with software maintainers. Our participants reflected a more nuanced concept of project/product management.

OSS System Admin: The system admin supports the base operational systems, selects, configures, connects, and fine-tunes the subsystems that are components of a robust and efficient larger part (Spinellis, 2006). This role includes activities related to providing technical support and system administration for OSS tools. A system admin provides tech support for OSS tools, and *"helps customers, walking them through registration, system setup, and all those different types of tasks that they need, they were trying to get done"*

(P4). This role is currently played by code contributors, although it is a well-defined role in many company-sponsored projects.

The interviews revealed multiple roles that spanned both technical and non-technical work. For example, we found a variety of community-centric roles, many of which were related to non-code contributions. We also found multiple individuals whose pathways were fluid, moving from code-centric roles to non-code-centric roles.

The subset of the roles and activities reported by our participants are similar to those in established software companies. These roles are increasingly sought out in OSS projects in addition to more traditional project-centric and code-related roles. The community-centric roles reported by our participants (including advocates, license managers, writers, strategists and mentors) are not usually well-recognized or easily identifiable from OSS project archives. Repository platforms (version control systems, issue trackers, project hosting sites) lack traces of these activities Carillo *et al.* (2017).

When analyzing our participants' pathways, we found that all of them were at some point paid to contribute to OSS by providing support and consulting—either by big companies or by working for their own companies. Even while being financially compensated, they mentioned that they performed parallel activities throughout their career, including volunteer (non-paid) activities in OSS projects.

Paid jobs were a starting point for an OSS career in some cases (P14, P4, P7). These participants mentioned working in the industry while simultaneously working on OSS-related activities at the same time (e.g., casually contributing, founding communities). Volunteering at OSS helped some participants start careers in closed-source projects, but even then many continued volunteering at OSS as a way to give back or maintain their reputation in the community.

4.3 Threats to Validity

In this section, we discuss the validity and reliability of our results for this study from the perspective proposed by Merriam Merriam and Tisdell (2015), as our analysis is qualitative.

Construct validity in qualitative research is related to the precise definition of constructs. To control this limitation, we compared and contrasted our construct definitions with the literature, with the interviewees' LinkedIn profiles, and OSS-renowned gray literature like Open Source Initiative (OSI) and Free Software Foundation (FSF) websites. During the coding phase, we used the constant comparison technique (Glaser and Strauss, 2017) whereby each interpretation and finding that emerges from the data analysis is compared with existing findings to increase construct validity.

Internal validity is related to the credibility that researchers were able to capture the reality as close as possible. One can argue that our interviewees all had successful OSS careers (and walked successful pathways) as they were speakers at an important OSS conference. Thus, our findings do not identify those pathways that ended in failure or disengagement. Moreover, a majority of our interviewees identified as women, which is a different distribution from typical OSS gender demographics (e.g. 11.2% Robles *et al.* (2016)). The roles uncovered in our study are not meant to be exhaustive, and further research into different ecosystems, domains, and types of contributors likely will uncover other roles and pathways. During our recruitment process, we reached out to OSCON invited speakers regardless of their gender. In our initial sample, out of the 11 interviewees, 6 identified as women. The sample reached via snowballing resulted in six other women (and no men). This can be due to two possible reasons. Firstly, it could be a self-selection

bias given that some participants knew previous work from the researchers and identified themselves with the study. Secondly, when recruiting via snowballing, the women interviewees recommended more women contributors who ended up participating. While these limitations exist, we minimized their effects by selecting interviewees with different backgrounds and expertise. Additionally, we compared our findings with the literature on software engineering and OSS to sharpen construct definitions and increase internal validity.

Reliability refers to the extent that the results can be replicated. In short, it is difficult to replicate qualitative research since human behaviors, feelings, and perceptions change over time. Thus, in the reliability thread, Merriam and Tisdell (2015) suggests checking the consistency of the results and inferences. According to Merriam and Tisdell (2015), consistency refers to ensuring that the results consistently follow from the data and there is no inference that cannot be supported after the data analysis. To increase consistency, we performed data analysis in pairs, which was consistently revised by two experienced researchers. We had weekly meetings to discuss and adjust codes and categories until we reached an agreement. In the meetings, we also checked the consistency of our interpretations, continually discussing our results based on the previous literature. We also performed member checking with five participants and they confirmed our interpretation with minor changes.

Finally, the results presented in this paper are related to Open Source communities, thus, we do not expect that all our findings will be applicable to other contexts. However, to allow replication of our study, we carefully describe our research method steps.

Theoretical saturation. A potential limitation in qualitative studies is not reaching theoretical saturation. The quality, rather than the size, of the sample of participants

is essential to increase our confidence in the results. In this study, we interviewed 17 participants with different perspectives and perceptions about the studied phenomenon. Our participants were diverse in terms of the number of years with OSS, experience in mentoring, roles, and highest academic qualifications. Further, these participants represented 15 different OSS projects that differ in size, domain, and ownership (company, community, and foundation). They were involved in many projects during their careers, changing roles within and across companies, as well as playing multiple roles in parallel. As mentioned previously, the number of interviewed participants was adequate to uncover and understand the core categories in any well-defined cultural domain or study of lived experience Bernard (2017). While we cannot claim saturation, our population helped us uncover a consistent and comprehensive account of the nature of OSS contributions, which included uncovering several novel roles and pathways.

4.4 Concluding Remarks

Roles in OSS projects, as reported by our participants, extend beyond the code-related, project-centric ones. Most of the women interviewed in our study perform community-centric roles, including as advocates, strategists, community managers, community founders, mentors, license managers, writers, and treasurers. While these roles do not produce code, they are important for the growth and sustenance of OSS, especially in the new OSS landscape. In addition, most of our interviewees (15 out of 17) evolved in their careers and at some point played community-centric roles, even when they had technical backgrounds and expertise in code-related skills.

While in this chapter we discussed the pathways of successful OSS contributors, in the next we take the perspective of the motivations that lead contributors to join and stay

in OSS.

Chapter 5

MOTIVATIONS TO CONTRIBUTE TO OSS PROJECTS

OSS currently enjoys a place of distinction in producing key technologies and providing learning; from the first study, we observed that OSS also offers different career opportunities. According to the literature, different genders can be driven to contribute by different motivations (Burnett *et al.*, 2010). Women can be motivated to contribute to OSS by both intrinsic and extrinsic motivations. For example, kinship represents a relevant intrinsic motivation for women to join OSS (Robles *et al.*, 2016), and payment had been shown to be an extrinsic motivation that affects more women than men (Prana *et al.*, 2021).

The results from this study were published in the 43rd International Conference on Software Engineering (ICSE) in 2021 (Gerosa *et al.*, 2021).

5.1 Method

The research design for this study is presented in Figure 5.1.

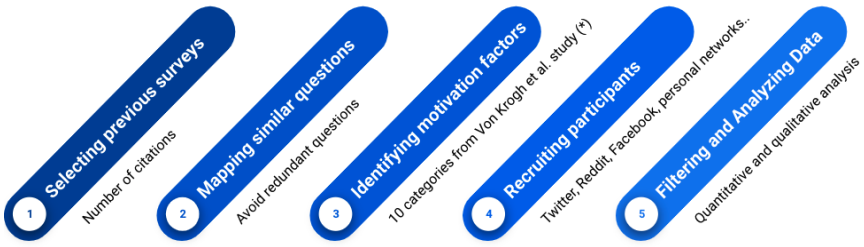

Figure 5.1: Research design for the study about OSS motivations

We collected data through a survey with 242 OSS contributors about the current mo-

tivations that drive OSS contributors to participate in OSS projects and selected prior surveys to help design our study. We started by searching for broad surveys about OSS motivations with a high number of citations on Google Scholar ((Lakhani and Wolf, 2003), (Hars and Ou, 2004)) and implemented another survey in the same period that collected a significantly higher number of respondents(Ghosh *et al.*, 2002)).

We used negotiated agreement Garrison *et al.* (2006) to group questions (or categories in the case of Hars and Ou (2004)) that could be considered similar and ended up with 20 questions extracted from the previous surveys.

To narrow down the analysis, we grouped the questions into higher-level constructs using the categories of OSS motivations from Von Krogh *et al.* (2012)'s study: *Ideology*, *Altruism*, *Kinship*, *Fun*, *Reputation*, *Reciprocity*, *Learning*, *Own-Use*, *Career*, and *Pay*.

After informed consent, we asked two open questions about motivation to contribute to OSS. The goal was to collect spontaneous answers before presenting participants with the list of motivation factors. To understand the shift in motivation, we asked participants why they first began and then continued contributing. On a new page, we presented the items from the previous step to identify the contributors' motivations on a 5-point Likert-scale.

We used the Likert-scale items to compare our results to the previous surveys. We ranked the questions based on the number of respondents who agreed to each motivation (checking "Agree" or "Strongly Agree"). We then compared each previous paper's ranking with a corresponding ranking of our answers, built by excluding the items that did not match those from the previous study. We analyzed the answers to the open questions that focused on motivation to start and to continue contributing. We categorized the answers based on a card sorting approach Spencer (2009).

5.2 Results

Figure 5.2 shows the answers to the Likert-scale items about what motivates OSS contributors, grouped per Von Krogh *et al.* (2012)'s categories.

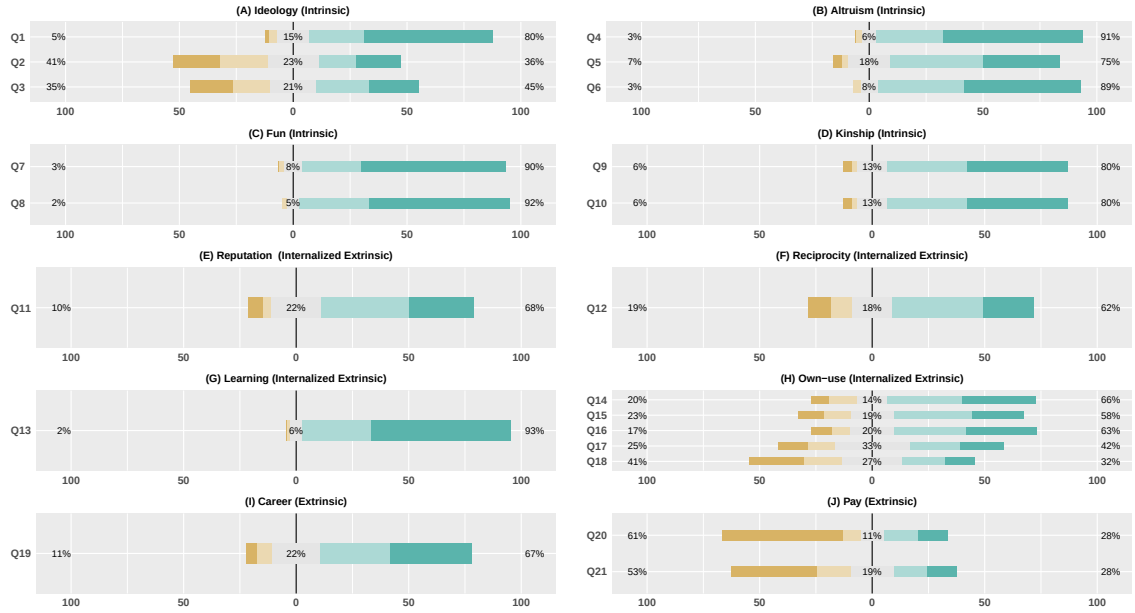

Figure 5.2: Responses to the 5-point Likert-scale items for motivation to contribute to OSS. The left hand (yellow) shows levels of disagreement, the middle (grey) shows neutral, and the right (green) shows levels of agreement.

Participants agreed that intrinsic motivations, especially *Fun*, *Altruism*, and *Kinship*, are key motivations—on average 91%, 85%, and 80% of the respondents agree (or strongly agree) that they contribute to OSS due to these. This is reflected in P164's excitement toward contributing to OSS: "*Discovered Linux and open source in general when I was a student in the 90s. Sending a patch across the ocean just seemed very exciting*," and P30's altruistic vision: "*To spread knowledge, which I think contributes to making a society better*".

Internalized-extrinsic motivations—*Learning* and *Reciprocity*—are also important fac-

tors. An impressive 93% of respondents (Figure 5.2(G)) agreed that they contribute because OSS allows them to learn and improve their skills, as P75 explained: "*I continued contributing to OSS projects because it was a good source for learning new things.*"

Extrinsic factors like *Career* and *Pay*, paint a contrasting picture. While 67% participants agree that OSS presents opportunity for professional growth (with only 11% disagreement, Figure5.2(I)), only 28% mention payment as a motivation (with 61% and 53% disagreeing with Q20 and Q21, respectively, Figure 5.2(J)).

Finally, some of the original motivations for contributing to OSS—*Ideology & Own-Use*—show mixed responses. Some aspects of ideology, such as opposing large companies and proprietary software, were not as popular as other motivations, which could be a result of large companies' recent embrace of OSS. On the other hand, the philosophy that source code should be open still remains strong (80%, Figure 5.2(A)). As P140 said "*I believe in the free software philosophy.*"

Own-Use is a mixed bag. "Scratch your own itch" was a key rallying call in the early days of OSS and is still a motivating factor for some, as P140 said: "*I contribute for my own purposes.*" However, the sentiment has changed. The two own-use questions with the biggest difference in opinions are Q18 and Q16 in Figure 5.2(H). Q18 relates to people seeking help from the community to realize their idea. While about 32% find this to be the case, a larger majority 61% show people joining existing communities. Interestingly, about 63% find proprietary software to be limited (Q16), at least in providing the same level of features, as P63 mentions: "*Depending on proprietary software was severely limiting in possibility, as with OSS we can fix our own bugs.*"

Results showed that social aspects, such as helping others, teamwork, and reputation have gained importance, while some intrinsic or internalized motivations are still preva-

lent, such as learning, fun, and altruism. Interestingly, OSS contributors often join because of extrinsic factors, but continue because of intrinsic factors. Although we had a small number of women respondents (18), we noticed that women more than men shift toward reciprocity in their motivation to stay in OSS. Only one woman started due to *Reciprocity*, but seven reported that they continued because of it; while this difference was more subtle for men (23 vs. 29). Moreover, we found that men are 4x more likely than women to report *Fun*, which corroborates the previous study from Burnett *et al.* (2010) that concludes women are generally more motivated to use technology to accomplish a goal rather than for fun.

5.3 Threats to Validity

We discuss some limitations and potential threats to the validity of the study.

Sampling bias. In our case, random sampling is not viable, since there is no single list of all OSS contributors. We combined multiple strategies to reach a broad and diverse sample, including a diverse population in terms of countries, projects, contribution roles, etc. Although the distribution of countries resembles the distribution of OSS contributions, there is a risk of a country bias. In terms of countries, USA (58) was dominant in North America (70), while Germany (23), UK (19), and Spain (18) are the most represented in Europe (100). We also have a low number of women and non-binary respondents, which mirrors our population's characteristic lack of diversity Bosu and Sultana (2019). Furthermore, we acknowledge that our sample may be biased in unknown ways, and our results are only valid for our respondents.

Response biases. As in any survey method, our work can have recall bias—respondents answer only what they recall and not necessarily what was most important to them in the

past. Recency and salience can also affect the respondents' answers. We aimed to reduce priming respondents with specific motivation factors by first presenting them with open questions, which allowed us to collect spontaneous answers.

Survivability bias. We focused our study on current OSS contributors. The motivations of those who tried but abandoned contributing may differ.

Self-selection bias. Participants decided whether they wanted to participate in the survey, and this may have influenced our results. Although most international OSS projects adopt English as their primary language, the language of the instrument may have influenced the participation of non-native speakers. Future studies might translate our survey and investigate regional differences.

Inappropriate participation. We employed several filtering and inspecting strategies to reduce the possibility of redundant participation and fake data; however, it is not possible to claim that our data is completely free of this threat.

Construct validity. To enhance construct validity, we based our survey on previous instruments. However, these instruments were not formally validated and may inadequately measure a given motivation. To mitigate this threat, we employed pilot studies to test and collect feedback about our instrument.

5.4 Concluding Remarks

The conventional wisdom of what attracts contributors to OSS does not apply to women. The most frequently cited motivation to start by women was *Own-Use*. While "scratch your own itch" has been seen as a typical motivation for joining OSS, the story is not so simple for women. Out of the six who mentioned starting due to Own-Use, only two respondents follow the Own-Use of starting on their OSS journey by finding a bug and

submitting a fix. The remaining four, while OSS consumers, had a different type of Own-Use. They started contributing through mentoring in structured programs (like GSoC) or personal contacts. This difference may be because the barriers to entry disproportionately disadvantage women Mendez *et al.* (2018a) or due to OSS's perception problem (e.g., toxic culture, code-centric). Another difference related to hedonic motives. While 20.4% (n=40) of the men stayed in OSS due to enjoyment and fun, only 5.5% (n=1) of the women reported these motivations to stay. Either way, OSS has a long way to go in attracting women.

OSS projects can leverage our results to devise and review strategies to support different genders to achieve their goals, resulting in more gender diversity in OSS communities. In the next chapter, we investigated a future perspective of goals through the perceptions of what it means to be a successful contributor.

Chapter 6

THE PERCEPTION OF BEING A SUCCESSFUL OSS CONTRIBUTOR

The way we define success has a remarkable impact on the choices we make in our personal and professional lives. Success in OSS encompasses more than code contributions alone. From the previous study (Trinkenreich *et al.*, 2020a), we found that some contributors perform a variety of non-code related activities (e.g., advocacy, technical writing, translation, project management) and follow different pathways than the celebrated "onion model" (Nafus, 2012; Trainer *et al.*, 2015; Trinkenreich *et al.*, 2020a). However, currently there is a misconception that success in OSS is only achieved through activities related to source code (Lakhani and Wolf, 2003; Fitzgerald, 2006; Robles *et al.*, 2019; Steinmacher *et al.*, 2017).

This study provided nuanced definitions of success perceptions in OSS, and show that OSS contributors have a broader perspective on success than the narrow focus on code-related activities—which is better supported by current tools and practices.

The results from this study were published in the Transactions of Software Engineering in 2021 (Trinkenreich *et al.*, 2021a).

6.1 Method

We investigated the self-definitions of success through interviews with 27 OSS contributors who are recognized as successful in their communities, and a follow-up open survey with 193 OSS contributors. The research design for this study is presented in Figure 6.1.

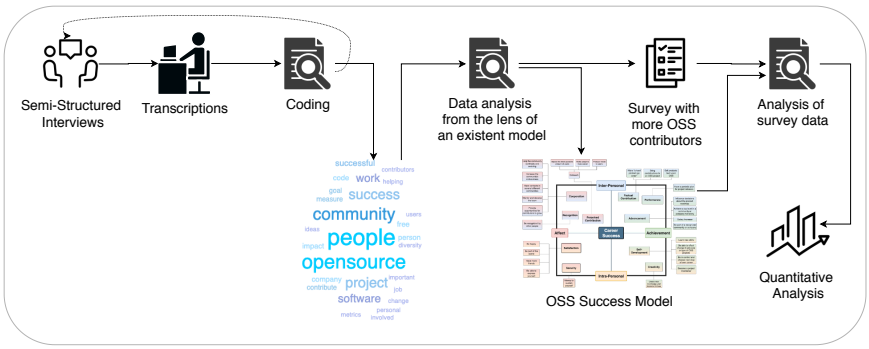

Figure 6.1: Research design for the study about success in OSS

The interviews revolved around the central question: *"How would you define being successful in Open Source?"* We approached this topic after establishing rapport with the interviewee, asking about their career story and contributions. We used a script to guide the different areas of inquiry, while also listening for unanticipated information during the flow of the conversation.

Our sample comprised paid and volunteer contributors across 20 different OSS projects (e.g., Kubernetes, Drupal, R, Noosfero, Fedora, Debian, GitLab), which vary in terms of the number of contributors (30 to 3,000 contributors), product domains (including infrastructure and user-application projects), and types (backed by foundations, communities, and companies).

We qualitatively analyzed the transcripts of the interviews by inductively applying open coding in groups, wherein we identified the definition of success that each participant provided. We built post-formed codes as the analysis progressed and associated them with respective parts of the transcribed text, so as to code the success definitions according to the participants' perspectives, who were identified as P1 to P27.

To organize our categories according to Dries *et al.* (2008)'s model, three of the authors conducted multiple card sorting sessions together (Spencer, 2009), arranging the codes

according to the regions of the model using continuous comparison (Strauss and Corbin, 2007) and negotiated agreement (Garrison *et al.*, 2006).

Next, we conducted an online survey to triangulate the interview results by gathering data from a different perspective (Easterbrook *et al.*, 2008) and a larger sample. In the survey, we asked two key questions about participants' perceptions of success and additional demographic-related questions, including the relationship with OSS (paid/unpaid), types of contributions, gender identity, country of residence, and age. The target population included any person who contributes to OSS.

We asked participants about their three main types of contributions and classified participants as "coder" if they selected "code developer" or "code reviewer" as one of the three main types of contributions. We classified as non-coders those who selected only a subset of these options: translation, documentation, mentorship, user support, community building, bug triaging, event presentations, advocacy and evangelism, creative work and design, and project management.

We used the categories from the interviews, classified into the regions of Dries *et al.* (2008)'s model, as the starting point of the qualitative analysis of the survey questions. We diligently analyzed the answers to identify any new perceptions of success that did not previously emerge from the interviews, but all survey answers could be mapped to the existing categories. We also used descriptive statistics to summarize the survey responses, their association with each other (success constructs), and the demographics data (Wohlin and Aurum, 2015).

6.2 Results

Our analysis of the interviews revealed 26 categories that explain how our participants defined success. We organized these categories using the multidimensional model of success proposed by Dries *et al.* (2008), as can be seen in Figure 6.2. The 26 categories covered all ten regions of the model. Table 6.1 presents the number of participants (interviews and surveys) whose responses fit in each region. The survey analysis did not provide any new definitions of success. In the following, we present our findings organized by quadrant.

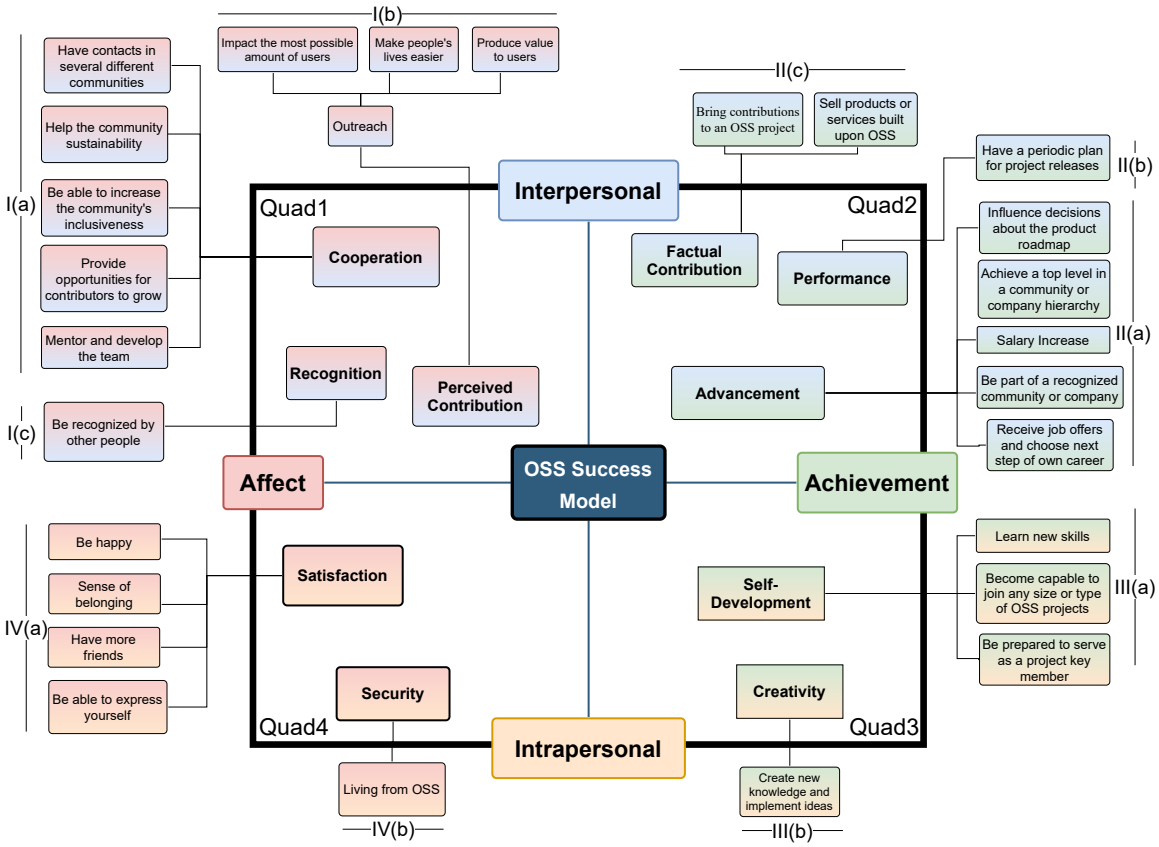

Figure 6.2: OSS Success Model. We mapped our participants' definitions (shown outside the bold square) to Dries *et al.* (2008)'s model, which organizes success in four quadrants.

Table 6.1: Success meanings from the interviews and the survey classified per Dries *et al.* (2008)'s model

	Region	Participants' IDs (Interviews)	# Interviews (total: 27)	# Survey (total: 193)
	Participants who mentioned at least one Region in Inter-personal		26	162 (84%)
Quad1	Cooperation	P2, P3, P4, P5, P7, P8, P13, P15, P17, P20, P21	11	15 (8%)
	Perceived Contribution	P5, P6, P9, P11, P17, P18, P21, P22, P25	9	57 (30%)
	Recognition	P1, P9, P13, P22, P23, P25, P27	7	29 (12%)
	Participants who mentioned at least one Region in Q1		**19**	**93 (48%)**
Quad2	Advancement	P1, P10, P12, P16, P20, P21, P22, P23, P24	9	8 (4%)
	Performance	P4	1	0 (0%)
	Factual Contribution	P2, P6, P12, P14, P18, P19, P20, P21	8	74 (38%)
	Participants who mentioned at least one Region in Q2		**15**	**81** (42%)
	Participants who mentioned at least one Region in Intra-personal		11	49 (25%)
Quad3	Self-Development	P7, P16, P18, P19, P20, P21, P24	7	19 (8%)
	Creativity	P10	1	2 (1%)
	Participants who mentioned at least one Region in Q3		**8**	**21** (11%)
Quad4	Satisfaction	P1, P5, P10, P16, P21, P26	6	14 (7%)
	Security	P19, P24	2	17 (7%)
	Participants who mentioned at least one Region in Q4		**7**	**30 (16%)**

Quad1: Interpersonal x Affect; **Quad2**: Interpersonal x Achievement; **Quad3**: Intrapersonal x Achievement; **Quad4**: Intrapersonal x Affect

The total per quadrant is not the sum of the regions since the participants often provided an answer that was categorized into **more than one region**.

Quad1: Interpersonal × Affect The first quadrant in Dries *et al.* (2008)'s model is defined by two dimensions: (1) interpersonal, which represents an individual's relationships with the outside world; and (2) affect, which represents internal feelings and perceptions that characterize success. This quadrant contains three distinct regions of meaning: cooperation, perceived contribution, and recognition.

Cooperation (Figure 6.2.I(a)) is defined as working with others (peers, superiors, subordinates, clients, etc.). The collaborative nature of OSS relates to this region as OSS contributors work together, support their community, and help their peers. In our analysis, we identified five categories, which we explain next.

Success included building social capital and networks of relationships, i.e., *"having contacts in several communities"*, as it allows quickly identifying sources of help when necessary (P8, P17, P21). It also includes being able to contribute to *"community sustainability"*, so it can be "as great as it can possibly be" (P3) and "more diverse and more inclusive" (P13). "Bringing people together" (P8) to *increase the community's inclusivity* was also repeatedly mentioned as a factor of success. Participants often mentioned individual success as part of the community's success: "having a healthy community is probably the most important thing" (P4) and "the sign of a healthy open source project is where everybody feels like their voice is heard and their opinion matters" (P7).

The cooperation aspect of OSS was also highlighted when participants defined their success as the ability to support others' success by *"providing opportunities for contributors to grow"* (P7) and "become more present and productive" (P15) by "giving everybody the opportunity [to climb] the contributor ladder" (P7).

Participants also cited success as being a *mentor* who is "friendly, didactic, and receptive to increase contributions" (P2 and P20), "who [neither] burn[s] themselves out, [nor acts as] the hero in the situation" (P15). An OSS mentor plays a crucial role in collaborative communities and influences the degree to which a newcomer relates to an OSS community and identifies with it (Carillo *et al.*, 2017). Indeed, our participants mentioned that newcomers need to "feel they are heard" (P3), and that successful mentors *develop the team* by "let[ting] people participate" (P4) and "being open to new ideas, whether that could be coding, helping to figure out what the roadmap is, identifying features, identifying bugs, kind of all those things coming together" (P4).

Perceived Contribution (Figure-6.2.I(b)), according to Dries *et al.* (2008), equates with serving society. In the context of OSS, our participants mentioned perceived con-

tribution from the perspective of *outreach*—i.e., "impact on people in the world" (P11). Participants considered themselves as successful when the product they contribute to has "high adoption"(P9), "produce[s] value for the people" (P17), and makes people's lives easier" (P5).

Recognition (Figure-6.2.I(c))—or being adequately rewarded and appreciated for one's efforts or talents (Dries *et al.*, 2008)—was also mentioned by our participants. P13, for example, defined success as *"being recognized by the community and the project's stakeholders."* P1 considered recognition as awareness that "the maintainer[s] of these projects know that they can [participate] as a subject matter expert" (P1).

Quad2: Interpersonal × Achievement As per Dries *et al.* (2008), this quadrant includes accomplishments external to the actor's self across three regions: advancement, performance, and factual contribution.

Advancement (Figure-6.2.II(a)) is defined as progressing and growing in terms of level and experience. In the OSS context, this relates to *influencing decisions* about the product, *"being part of an influential community that is well recognized, a community that you say the name and people know what is*" (P21), *receiving job offers*, "writing [one's] own ticket" in one's career (P12), receiving a *salary increase*, or *achieving a top-level position.* "Money" in some cases represented growth (e.g., "salary going up" (P16)), which differs from some other cases in which money represented a way to earn a living from OSS, which we classify as Security.

The Performance (Figure 6.2.I(b)) region is defined as attaining verifiable results and meeting set goals (Dries *et al.*, 2008). In our context, this translated to having a *plan for project releases* "depending on what the goals of the project are, such as working on a new

release every six months" (P4). Project planning activities demonstrated the relation of the actor to the external world (interpersonal dimension), as explained by P4: "if [one is] not making [the release], [they are] letting a lot of people down".

Factual contribution (Figure-6.2.II(c)) is about individual contributions to the collective (Dries *et al.*, 2008). An indication of success in this region includes *bringing contributions to an OSS project*, by "getting a change that you wrote accepted" (P12), including "a code change, a documentation change… [or otherwise] getting something you made merged" (P12). Besides code contributions, interviewees mentioned implementing ideas or any type of revisions or contributions to the project, as well as "actively reviewing and looking at what people are suggesting" (P2). Contributions can also represent something tangible, such as achieving financial gains when "selling the platform" (P6) or when having a "ventured organization" (P6).

Quad3: Intrapersonal × Achievement Dries *et al.* (2008) describe this quadrant as including real accomplishments of the actor's "self." It contains two distinct regions of meaning: self-development and creativity.

Self-Development (Figure-6.2.III(a)) is defined as realizing one's potential through self-management of challenges and learning experiences (Dries *et al.*, 2008). This has been a classic motivation for contributing to OSS (Hertel *et al.*, 2003; Hars and Ou, 2004; Lakhani and Wolf, 2003). However, success definitions mentioned by the interviewees go beyond *"learning new skills"* (P16). They also include the path to receive a promotion, as stated by P20: "I reviewed other people's code to improve my review skills to become a maintainer," and *be prepared to serve as a key project member* by "being a mature reviewer and contributor" (P2) *"capable of effecting change in an open source project, from the small to*

the large" (P7).

Creativity (Figure-6.2.III(b)) is about making something innovative and extraordinary (Dries *et al.*, 2008). We found this to mean the freedom to *"create new knowledge"* (P3), but also *"propagat[e] ideas"* (P3). Creativity is relevant to the OSS context as individuals from innovative communities have greater opportunities to express themselves and experience a sense of accomplishment Lakhani and Wolf (2003).

Quad4: Intrapersonal × Affect The intrapersonal × affect quadrant includes feelings and perceptions that characterize the career of an actor's "self" (Dries *et al.*, 2008), which contains two regions: satisfaction and security.

Satisfaction (Figure-6.2.IV(a)) is about achieving happiness and personal satisfaction, either in the family or work domain (Dries *et al.*, 2008). Participants mentioned satisfaction as *"being happy"* (P1, P16, P26), which also included *"being able to express yourself"* (P10). They talked about their *sense of belonging* and "need for emotional inclusion" (P16), the importance of "participating in the world that is being created" (P10), and having *"a ton of friends* and people [to] hang out with or chat with, about nontechnical stuff" (P5).

Security (Figure-6.2.IV(b)) means meeting one's financial or employment needs (Dries *et al.*, 2008). Participants characterized success as the ability to *make a living from OSS*—to "receive money as an OSS developer" (P24) and "prioritize what [financially] sustains you" (P19).

Survey analysis We conducted a survey to triangulate the definitions of success we identified from the interviews, expanding our population and exploring whether we could find any new definitions of success. We qualitatively analyzed the 193 answers to our

survey open question. Similar to interviews, the participants often provided multiple definitions, which could be categorized into more than one region from Dries *et al.* (2008)'s model. However, no new category emerged from the survey analysis.

We looked deeper into the survey results to understand the prevailing definitions of success among our respondents and across different demographics. When presenting the results, we use supplementary and corroborative counting of the survey responses to triangulate the qualitative analysis of the definitions of success Hannah and Lautsch (2011).

The dimensions of success. The majority of respondents defined success in terms of a relationship with the external world (interpersonal) rather than the actor's self (intrapersonal), accounting for 84% vs. 25% of respondents. For the interpersonal dimension, respondents identified success across both ends of the affect and achievement spectrum—25% were related to the affect dimension and 49% were related to achievement. When considering definitions related to the intrapersonal dimension, none of the regions were mentioned by more than 10% of the respondents. This preponderance of definitions related to the interpersonal side could be due to the collaborative nature of peer-production sites such as OSS, where contributing to a common good and being recognized for it have been cited as key motivation factors (Von Krogh *et al.*, 2012; Gerosa *et al.*, 2021; Hertel *et al.*, 2003; Hars and Ou, 2004; Roberts *et al.*, 2006).

In fact, Factual (38%) and Perceived Contribution (30%) were the most mentioned regions, followed by Recognition (12%). None of the other regions across all quadrants had more than 10% of responses. These responses reflect that, in OSS, while contributions matter, the way that others (community, peers, society) value the contributions is also an important indicator of success.

Respondents who identified Factual Contribution as a definition of success em-

phasized that the number, size, and frequency of contributions can be objective concepts to quantify a significant contribution to the community. They defined success as *"finding a way to sustainably contribute"* (S25), or being *"someone who is able to regularly contribute"* (S11) and *"spending time on the project often"* (S68). A successful contributor is one who provides *"a wide spectrum of contributions"* (S6). Moreover, respondents identified various types of contributions across different project-centric or community-centric roles (Trinkenreich *et al.*, 2020a), as mentioned by S2: *"Successful contributors add or change major features, and organize the community"*.

Those who considered Perceived Contribution as success emphasized the importance of their contribution, such as publishing and maintaining software that is used by and useful to a lot of people. According to S136, the perceived value of their contribution could be measured by *"how many people have used the OSS code and how much value has it created"*. Some of these definitions of success in OSS included: *"someone who publishes and maintains software that is useful for a lot of people or for the user community"* (S3) and *"when the software solves and helps real-world problems"* (S169).

Finally, our respondents reflected many different perceptions of success related to Recognition in their community; which included *"having a high number of stars on the own repository in GitHub"* (S58 and S109), *"receiving donations"* (S21), and *"being invited for conference invites/talks"* (S16).

Demographics and the meaning of success As recent literature has shown, the OSS community is becoming more diverse in terms of the gender of contributors, types of contributions, and financial rewards (Carillo *et al.*, 2017; Trinkenreich *et al.*, 2020a). We took a deeper look into these demographic subgroups with respect to their definitions of success. Understanding how different demographics perceive success can sup-

port creating mechanisms to better support diverse contributors and improve the state of diversity in OSS. Figure 6.3 illustrates the definition of success for each demographic subgroup. The percentages in the figure reflect the number of participants who mentioned any meaning under each quadrant per subgroup. For example, 80 participants who identified themselves as code contributors reported at least one meaning of success categorized in Quad1. Therefore, given there were 163 code contributors, 49.1% of the code contributors in our sample associated success with cooperation, perceived contribution, or recognition (Quad1).

From the 193 survey respondents, 165 identified as men, 16 as women, and 2 as non-binary. The gender distribution of our respondents matches that of those reported in other OSS studies ((Vasilescu *et al.*, 2015b; Robles *et al.*, 2016; Singh, 2019b)). We dropped from this analysis the 10 respondents who did not disclose their gender.

Although not having statistical difference between genders, we found that women more than men from our sample include recognition in their definition of success. The literature shows that men relate success to tangible and objective outcomes, but, contrary to the research in other domains (Dyke and Murphy, 2006; Cho *et al.*, 2017; Porter, 2019), definitions of success that are considered subjective were also cited by men. When looking at the dimensions, both men and women more frequently mentioned success definitions classified in the interpersonal quadrants (Quad1/Quad2) than those in intrapersonal (Quad3/Quad4).

Our survey included answers from 163 *coders* and 30 *non-coders*, i.e., those who work only on non-code related activities (e.g., advocacy, license management, technical writing). We could not find statistically significant differences between the distribution of answers from the two subgroups. We could also not find statistically significant differ-

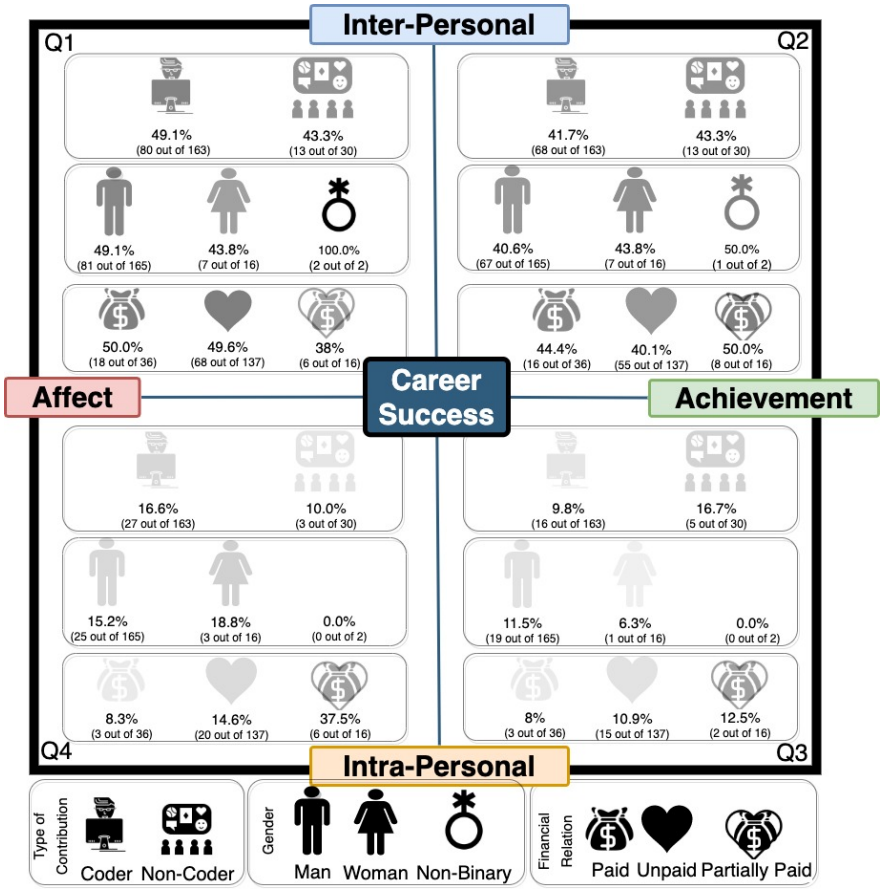

Figure 6.3: Subgroup analysis of the meanings of success. The opacity of the icons represents the percentage of each group in the quadrant. Darker means a higher and lighter a lower percentage. Some respondents provided answers about success that accounted for more than one quadrant.

ences when sub-grouping based on compensation (paid vs. unpaid). The statistical test results including the p-values of these comparisons are in the supplementary material.

6.3 Threats to Validity

In the following, we discuss the validity and reliability of our results from the perspective proposed by Merriam and Tisdell (2015).

Internal validity. The characteristics of our sample may have influenced our results. A great part of our interviewees (11 out of 27) were speakers at an OSS conference and half (13 out of 27) of the interviewees identified as women, even though we did not push for an equal gender split. This diversity of profiles helped bring a more diverse perspective on the phenomenon. Our survey, which received almost 200 answers, corroborated our results. The distribution of our survey demographics is similar to the larger OSS population as reported elsewhere (Zlotnick, 2017b; Robles *et al.*, 2016; Vasilescu *et al.*, 2015b).

Survival bias. Our results reflect the opinion of current contributors who joined OSS and made it past the initial contribution barriers (Steinmacher *et al.*, 2015b). Therefore, to promote diversity in OSS, we acknowledge that additional research is necessary to understand success from the perspective of both those who do not make it past the initial barriers and those who are currently not attracted to OSS.

Recall bias. Moreover, as our survey question was open-ended, our results could be impacted by either salience bias, where respondents focus on definitions that are prominent or emotionally striking and not necessarily all the factors that matter; or by memory bias, where participants answer questions based on what they can first recall and not necessarily what is most important to them.

Data Consistency. Consistency refers to ensuring that the results consistently follow from the data and there is no inference that cannot be supported after data analysis (Merriam and Tisdell, 2015). The same group of researchers performed the qualitative analysis of interview transcripts and survey responses. We had weekly meetings to discuss and adjust codes and categories until reaching an agreement. In the meetings, we also checked the consistency of our interpretations, continually discussing our results based on definitions of Dries et al.'s model (Dries *et al.*, 2008). All analysis was thoroughly grounded in

the data collected and exhaustively discussed among all team members, which includes researchers with extensive experience in qualitative methods.

Theoretical saturation. A potential limitation in qualitative studies regards reaching theoretical saturation. In this study, we interviewed 27 participants with different backgrounds and perceptions about the studied phenomenon. The participants are diverse in terms of gender, number of years involved with OSS, and highest achieved academic degree. We continued inviting participants until we could not find any new concepts for five consecutive interviews. Moreover, we collected answers from 193 respondents about what it means to be a successful OSS contributor, and did not find any new meanings. Therefore, although theoretical saturation cannot be claimed, we believe that we obtained a consistent and comprehensive account of the phenomenon.

6.4 Concluding Remarks

Success is a multifaceted and complex concept, including both objective metrics and subjective perceptions of accomplishments. The interpersonal dimension plays a dominant role in the definition of success, in which factual and perceived contributions are the most referenced, followed by recognition. Although we did not find a statistical difference in the perception of success across genders, most of the women reported interpersonal perceptions of success, equally in the affect (subjective) and achievement (objective) dimensions. When looking at the interpersonal dimension, cooperation was not mentioned by any women, and recognition was relatively more cited by women (25.0%) than by men (14.6%). In the next chapter, we investigate different factors in a large and community-oriented OSS project.

Chapter 7

THE LINUX KERNEL CASE STUDY: UNDERSTANDING THE SENSE OF VIRTUAL COMMUNITY AND CHALLENGES

Given the diversity of characteristics of different OSS projects, we decided to conduct a more focused set of studies to understand specific points about the involvement of women in OSS. We studied one specific community to avoid confounding factors related to differences that each OSS community can pose. Introduced in 1991, Linux Kernel represents one of the largest and most active OSS projects (Homscheid, 2020), boasting over ten million source lines of code and more than 12,000 contributors from different countries and cultural backgrounds, including volunteers and paid developers from more than 200 companies (Tan *et al.*, 2020). While the Linux Kernel Mailing List is known for its uncivil comments and toxic discussions that can discourage people from joining the community (Miller *et al.*, 2022), community leaders aim to change the project's contentious image and increase the sense of community among members.

The Linux Kernel has a group of community managers to understand the state of diversity and inclusion, who closely collaborated with us in co-designing the data collection instrument and reaching out to potential participants. The study performed in Linux Kernel had the goal to explore factors that could potentially lead to a non-inclusive environment

or that could be harming the retention of contributors.

7.1 Study Design

We administered an online questionnaire using LimeSurvey, a leading Open Source survey software, to survey Linux Kernel contributors. In the following, we discuss our approach and instrumentation.

7.1.1 Planning the measurement instrument

The questions were discussed during 12 online meetings between October 2020 and February 2021 with a group of five researchers experienced in both OSS and survey studies and two Linux Kernel community managers. The group discussed each of the questions until reaching consensus.

We used measurement instruments from prior literature where possible. Considering the complexity of a person's decision to participate in a project or not, we included questions to explore forces that push contributors (i) towards or (ii) away from a project (Steinmacher *et al.*, 2014d). Investigating the forces that impact people with different individual characteristics can help us better support a diverse community Gerosa *et al.* (2021). **(i) Attractiveness forces**: The questions about what drives contributors toward projects included one Likert question about motivations to contribute and six Likert questions about the sense of virtual community. The question about motivations to contribute was based on Gerosa *et al.* (2021)'s instrument, which was built upon previous studies of motivations in OSS (Lakhani and Wolf, 2003; Ghosh *et al.*, 2002; Hars and Ou, 2004), as we present in Table 7.1. Following the community managers' request to make the questionnaire as short as possible, we grouped the motivation factors from Von Krogh *et al.* (2012)'s

study into three factors: 1. social motives (kinship and altruism) (Neel *et al.*, 2016); 2. hedonic motives (joy and fun) (Tamilmani *et al.*, 2019); 3. moral motives (ideology and reciprocity) (Janoff-Bulman and Carnes, 2018); and 4. extrinsic motives (career and pay).

Table 7.1: The grouping of motivation factors

Grouped factors used in our study	**Von Krogh *et al.* (2012)'s factors**
Hedonic Motives	Enjoyment and Fun
Moral Motives	Reciprocity and Ideology
Social Motives	Altruism and Kinship
Extrinsic Motives	Pay, Career, Reputation, and Learn

The six questions about feelings of a virtual community were adapted from both (Good *et al.*, 2012)'s Sense of Belonging instrument and Blanchard *et al.* (2011)'s Sense of Virtual Community (SVC) instrument, to better fit with the context of OSS contributions, as we present in Table 7.2. In collaboration with a group of Linux Kernel community managers, we analyzed the items proposed by Blanchard *et al.* (2011) and decided to use a subset of questions to compose a shorter version of the instrument to cover the dimensions of SVC. The subset was synchronously discussed by researchers and managers, and the items were considered appropriate and meaningful to represent SVC to the Linux Kernel contributors. **(ii) Veering forces**: The question about difficulties contributing was created as an open question to allow participants to define their ideas prior to researcher guidance. We aimed to collect data that could be inductively analyzed by inquiring about the challenges that were specifically happening in the Linux Kernel community (Bernard and Gravlee, 2014).

The questionnaire started with an informed consent, followed by closed questions about the importance of each motivation factor as a reason to contribute to the Linux Kernel, six questions about their feelings about the Linux Kernel community, and one

Table 7.2: The scales to measure the contributors' sense of virtual community

	Our Scales	**Original Scales**
svc1	I do not feel at home in the group	I feel at home in this group (Blanchard, 2007)
svc2	I feel that I belong to the group	I feel that I belong to the math community (Good *et al.*, 2012)
svc3	If I have a problem, I know members in the group who I can ask for help	If there is a problem in this group, there are members here who can solve it (Blanchard *et al.*, 2011)
svc4	I want to contribute more but I do not feel valued	I feel valued (Blanchard *et al.*, 2011)
svc5	A majority of developers in the group know me	Very few other group members know me (Blanchard *et al.*, 2011)
svc6	The majority of the developers and I want the same thing	Other members and I want the same thing from this group (Blanchard *et al.*, 2011)

open question about the challenges faced while contributing to the Linux Kernel. Finally, we added demographic questions aiming to segment analysis and understand the phenomenon considering the different dimensions of our participants, and an open question for additional comments. The demographic questions included gender identity, English confidence, financial compensation, starting year at Linux Kernel, and country of residence. The complete instrument can be found in the online appendix (Trinkenreich, 2022).

7.1.2 Piloting the questionnaire

We piloted the questionnaire in two rounds. In the first round, we sent the questionnaire to five Linux Kernel maintainers, who answered it and provided feedback via email. At this point in the piloting process, we received positive feedback on the survey design; one of the respondents replied to the community manager: *"Thanks for sharing. The questions are well-designed and meaningful. It was a pleasure taking the survey."* No further adjustments were made. In the pilot study, maintainers suggested to reverse-code

some items for the SVC construct, i.e., items worded as negative statements (low score indicates agreement). Indeed, inverse, negative, or reverse-coded items can be defined as those having a directionality opposed to the logic of the construct being measured Weijters and Baumgartner (2012). Using negative items can help mitigate acquiescence bias Baumgartner and Steenkamp (2001) that can occur when participants tend to agree with statements without regard for their actual content, or due to laziness, indifference, or automatic accommodation to a response pattern Podsakoff *et al.* (2003). We inverted two of the four items. The item *I feel at home in the group* was changed to *I don't feel at home in the group.* We inverted and adapted the question *I feel that my contribution is valued* to *I want to contribute more but I do not feel valued.*

After the first pilot, we revisited the questionnaire and ran two more pilot sessions with two researchers who are open source contributors. We used the think-aloud method Van Someren *et al.* (1994) and recorded their suggestions while answering the questions. We made minor changes to the questionnaire and increased font size for better readability on different devices.

7.1.3 Recruiting participants

The community managers who worked in collaboration with the researchers recruited participants by sending emails to the different mailing lists in the Linux Kernel community. Further, we presented the study goals and the concepts explored in the survey during the first day of the Linux Plumbers annual conference (https://lpc.events/event/11/), inviting participants to answer the questionnaire. The survey was available between August 12 and September 21, 2021.

We received 316 answers. No questions were mandatory, so not all categories sum to

316. The demographics are presented in Table 7.3.

Table 7.3: Demographics of the Linux Kernel respondents (n=316)

Attribute	N	Percentage
Gender		
Man	262	82.9%
Woman	33	10.4%
Non-binary	6	1.9%
Prefer not to say	10	3.3%
Prefer to self describe	3	0.9%
Blanks	2	0.6
Continent of Residence		
Europe	147	46.5%
North America	83	26.3%
Asia	66	20.9%
South America	8	2.5%
Oceania	4	1.3%
Africa	2	0.6%
Blanks	6	1.9%
Starting year at the Linux Kernel		
2000 or earlier	33	10.4%
Between 2001 and 2010	96	30.4%
Between 2011 and 2021	184	58.2%
Blanks	3	0.9%
Current Compensation for the Linux Kernel contributions		
Paid	177	56.0%
Unpaid (volunteer)	139	44.0%

We used the data collected through this survey in different studies that are detailed in their sections and introduced below:

- Attractiveness forces: conception and Evaluation of a theoretical model of Sense

of Virtual Community (SVC) in Linux Kernel. For this specific study, we used the answers to questions about motivations to join Linux Kernel, SVC, English confidence, and demographics (gender, tenure, country of residence, and information about payment to contribute). The details of this study are presented in detail in Section 7.2.

- Veering forces: The challenges faced by Linux Kernel contributors. For this study, we used questions about challenges faced by the contributors and information about their gender, tenure and information about payment to contribute). The data analysis and results are presented in detail in Section 7.3.

7.2 A Theoretical Model of Sense of Virtual Community in Linux Kernel

Hagerty *et al.* (1992) defined a sense of belonging as *"the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment."* The need to belong is a powerful, fundamental, and pervasive force that has multiple strong effects on emotional patterns and cognitive processes, across all cultures and different types of people (Baumeister and Leary, 2017). Maslow (1943) positioned 'belonging' as a basic human need, and Hagerty and Patusky (1995) posited that a sense of belonging represents a unique mental health concept. A sense of belonging is key to productivity, satisfaction, and engagement (Baumeister and Leary, 2017), and can help to avoid attrition (Allen, 2019). In Science, Technology, Engineering, and Mathematics (STEM), a sense of belonging is strongly related to retention (Espinosa, 2011), especially for underrepresented groups (Happe and Buhnova, 2021).

The sense of belonging that members have towards others within a certain group is known as a *sense of community* (Burroughs and Eby, 1998). The dimensions of a sense of

community include feelings of membership and attachment to a group (Blanchard, 2007), and a feeling that members matter to one another and to the group (McMillan and Chavis, 1986). The concept of a sense of virtual community (SVC) was developed by observing that virtual communities represent a new form of community, in which social relationships are predominantly forged in cyberspace (Koh *et al.*, 2003). Experiencing a sense of belonging is significantly related to STEM retention (Espinosa, 2011). Belonging is a top motivation and even more relevant for underrepresented groups (Espinosa, 2011; Johnson, 2012; Thoman *et al.*, 2014). Women from STEM colleges tend to have a lower sense of belonging than their counterparts, face additional challenges to achieve success, and are more likely to quit (Beyer *et al.*, 2004; Cheryan *et al.*, 2009; Hurtado and Carter, 1997; Johnson *et al.*, 2007; Strayhorn, 2012).

Understanding SVC in OSS is relevant as it can influence the vitality and sustainability of a community (Blanchard, 2008; Tonteri *et al.*, 2011), and is linked to more satisfied, involved, and committed contributors (Kim *et al.*, 2020). Individuals who develop a psychological and relational contract with a community are focused on that feeling of connection, rather than external factors such as earning something or climbing a career ladder, and therefore tend to develop a deeper, reciprocal relationship with that community (Burroughs and Eby, 1998).

In the OSS context, in a previous study (Section 5) we found that contributors often shift from extrinsic motivations (when they join) to intrinsic ones (as a reason to stay). Since the reasons to continue shift to social factors, contributors might experience an increase their feelings of belonging and sense of virtual community (Sax *et al.*, 2018). Supporting this, previous research showed that social capital supports long-term engagement for both men and women in OSS projects Qiu *et al.* (2019c).

The results from this study were accepted by the 45th International Conference on Software Engineering (ICSE) 2023 (Trinkenreich *et al.*, 2023).

7.2.1 Method

We developed a theoretical model of SVC grounded in prior literature. We then evaluated our model through a sample (N=225) of Linux Kernel project contributors, using Structural Equation Modeling. The results of our analysis provide empirical support for part of our model, showing that *hedonism* (motivation that aims to maximize pleasure and fun and minimize pain (Tamilmani *et al.*, 2019)) and *social motives* (motivation that aims to maximize collaboration and others' gains (McClintock, 1972) have a positive association with a sense of virtual community, which can be weakened when contributors are *being paid* or are surrounded by an *authoritative culture*). The model was built for all genders, and we evaluated how gender, tenure, and English confidence affect SVC. We found that identifying as part of gender minority groups (women, non-binary, and self-describing own gender) and having lower self-confidence in English proficiency tend to be factors that reduce the sense of virtual community.

We used Partial Least Squares–Structural Equation Modeling (PLS-SEM) to analyze the relationships that happen between motivations (Ringle *et al.*, 2015) and a sense of virtual community. SEM is a second-generation multivariate data analysis method; a recent survey (which also provides an introduction to the method) indicates that PLS-SEM has been used to study a variety of phenomena in software engineering (Russo and Stol, 2021). SEM facilitates the simultaneous analysis of relationships among constructs, each measured by one or more indicator variables. The main advantage of SEM is the ability to measure complex model relationships while accounting for measurement error when

using latent variables (e.g., sense of virtual community). PLS-SEM has previously been used in literature to evaluate factors that impact the sense of belonging in other contexts (Chen and Lin, 2014; Ellonen *et al.*, 2013).

The research design is summarized in Fig. 7.1.

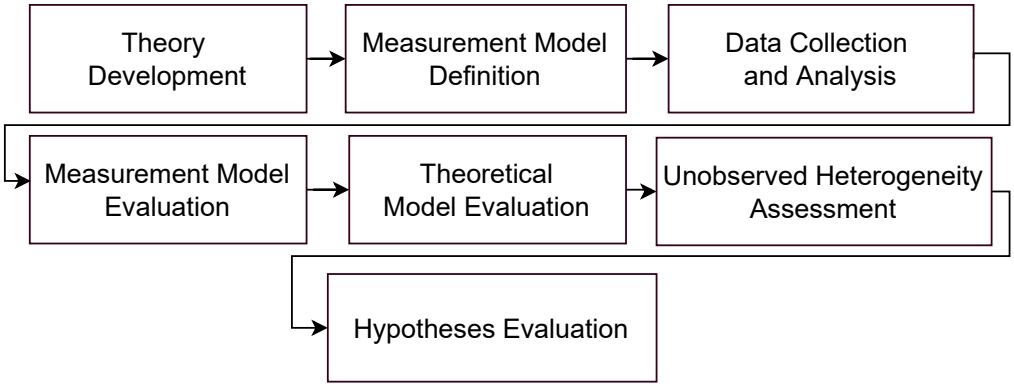

Figure 7.1: Research Design and Phases for Results' Analysis

Theory Development Figure 7.2 presents a graphical overview of the proposed theoretical model.

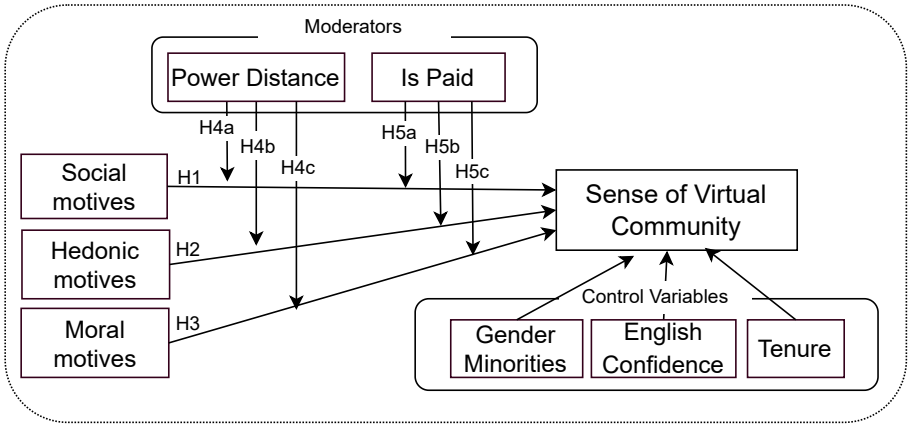

Figure 7.2: Research model

Feelings of belonging in an online community are influenced by a number of individual characteristics and factors of the surrounding environment (Allen, 2020). In the

education literature, researchers (Goodenow and Grady, 1993; Solomon *et al.*, 1996) found associations between students' sense of belonging and a range of motivational variables. Motivational factors can be regarded as expectations prior to or during interaction with a virtual community (answering *why* users behave). Integration and fulfillment of needs refer to the idea that common needs, goals, and beliefs provide an integrative force for a cohesive community that can meet both collective and individual needs. Thus, meeting members' needs is a primary function of a strong community (McMillan and Chavis, 1986).

The role of social motives Individuals who develop a psychological relationship contract with a community because it is focused on a state of being involved tend to develop a sense of community (Burroughs and Eby, 1998). Previous research on online communities also showed that individuals who are driven by *social motives* Neel *et al.* (2016) tend to develop a sense of virtual community Kim and Drumwright (2016); Chang *et al.* (2016). Based on the Fundamental Social Motives Inventory, we included both kinship and altruism as social motives Neel *et al.* (2016), and propose the following hypothesis:

Hypothesis 1 (H1). Open Source contributors driven by social motives have a higher sense of virtual community.

The role of hedonic motives Most of the respondents in Gerosa *et al.* (2021)'s study (91%) agreed (or strongly agreed) that they contribute to OSS for entertainment (fun). Hedonic motivation is a type of motivation that aims to maximize pleasure and fun and minimize pain; it is an umbrella term that includes hedonic expectancy, perceived enjoyment, and playfulness (Tamilmani *et al.*, 2019). Considering that expectations of enjoyable

experiences, feelings of amusement, and being mentally or intellectually stimulated by interactions are associated with a sense of virtual community (Koh *et al.*, 2003; Tonteri *et al.*, 2011), and that changes in the perceived fulfillment of entertainment needs can cause a change in sense of virtual community (Sutanto *et al.*, 2011), we propose the following hypothesis:

Hypothesis 2 (H2). Open Source contributors driven by hedonic motives have a higher sense of virtual community.

The role of moral motives It is known that some open source contributors have a strong ideological basis for their actions Stewart and Gosain (2006), believing, for example, that source code should be freely available. Recently, however, a study showed that ideology is not a popular motivation, especially for young contributors Gerosa *et al.* (2021).

Historically, the group-based morality of 'fighting' a shared dominant opponent incites a sense of virtual community among contributors (McGowan, 2001)—the classic example is Microsoft and its proprietary software in the 1990s, and its characterization of open source as 'communism,' and Linux as a 'cancer.' Besides ideology, we include reciprocity in moral motives, as it represents the moral desire of contributors who aim for social justice by giving back to the community (Janoff-Bulman and Carnes, 2018).

According to the Social Identity theory (Tajfel and Turner, 2004), sharing a moral vision is positively associated with feelings of belonging. Moreover, a homogeneous ideology throughout a religion was shown as being positively associated with a sense of virtual community (Gan *et al.*, 2019). Hence, we posit that:

Hypothesis 3 (H3). Open Source contributors driven by moral motives have a higher sense of virtual community.

The moderation role of power distance Motivations may not always be strong enough to sustain an OSS contributor's participation (Fang and Neufeld, 2009). Motivations may vary for different groups of people, depending on contextual factors. This implies the existence of moderating factors that change the relationship between motivations and a sense of virtual community. Cognitive Theory suggests that feelings of autonomy are positively associated with intrinsic motivations and belonging, while tangible rewards introduce a negative effect on intrinsic motivating factors (Deci and Ryan, 1985).

We evaluated the role of a feeling of autonomy using the variable of *power distance* from Hofstede (2001)'s framework of country culture as a proxy; a lower power distance would reflect in higher autonomy. We also evaluated the exposure to tangible rewards using the variable *is paid*.

People in societies exhibiting a large degree of power distance accept a hierarchical order in which everybody has a place and which needs no further justification (Hofstede, 2011). In high power distance cultures (where the existence of a high power differential between individuals is accepted and considered normal), information flows are usually constrained by hierarchy (Hofstede, 2001). As an important cultural value describing the acquiescent acceptance of authority, power distance has received increasing attention in many domains (Fock *et al.*, 2013; Auh *et al.*, 2016).

Prior research showed that, when living in cultures with a high degree of power distance, students reported a lower sense of belonging to their school (Cortina *et al.*, 2017). Leaders in hierarchical cultures need control over the information flow, and the desire to restrict autonomy and access to critical information by lower-level members of the team could lead to significant organizational barriers to sharing knowledge and working in the community (Ardichvili, 2008). Thus, we define the following moderation hypotheses:

Hypothesis 4a (H4a). Power distance moderates the association between Open Source contributors' social motives and their sense of virtual community.

Hypothesis 4b (H4b). Power distance moderates the association between Open Source contributors' hedonic motives and their sense of virtual community.

Hypothesis 4c (H4c). Power distance moderates the association between Open Source contributors' moral motives and their sense of virtual community.

The moderation role of compensation The traditional notion of OSS developers as volunteers is now long outdated, as many OSS contributors today are paid, employed by a company and tasked to contribute (Schaarschmidt and Stol, 2018; Taylor and Dantu, 2022; Trinkenreich *et al.*, 2020b). For over a decade, a significant number of Linux Kernel contributors have been paid to make their contributions, compensated by firms that have business models relying on the Linux Kernel (Corbet *et al.*, 2012; Riehle *et al.*, 2014; Homscheid *et al.*, 2015).

In contrast to traditional paid software development work, and despite its benefits to OSS contributors, introducing financial incentives in OSS communities create complex feelings among OSS developers (Sharma *et al.*, 2022). Developers on the Debian project, for example, expressed negative emotion because they felt payment went against the project's espoused values (Gerlach *et al.*, 2016). On the other side, not receiving pay for their work to support their livelihoods can frustrate OSS developers and affect their contributions (Sharma *et al.*, 2022).

In spite of compensation, OSS contributors may be driven towards a project by both simultaneous feelings of belonging (intrinsic) and payment (extrinsic) (Roberts *et al.*, 2006; Schaarschmidt and Stol, 2018). In any case, there is no research examining the complex

impact of receiving payment on intrinsic factors associated with SVC. As so many OSS developers today are paid, we would expect that the behavior of those who are paid and those who are not (volunteers) would diverge. Hence, we propose the following three moderating hypotheses:

Hypothesis 5a (H5a). Being paid moderates the association between Open Source contributors' social motives and their sense of virtual community.

Hypothesis 5b (H5b). Being paid moderates the association between Open Source contributors' hedonic motives and their sense of virtual community.

Hypothesis 5c (H5c). Being paid moderates the association between Open Source contributors' moral motives and their sense of virtual community.

To evaluate our theoretical model, we conducted a survey, targeting Linux Kernel contributors, as introduced in 7. In the following, we discuss the measurement model (i.e., operationalization of constructs), data collection, and analysis.

Designing the measurement model The theoretical model comprising the hypotheses is based on a number of theoretical concepts; some of the concepts may be directly observed (e.g. 'is paid'), but others cannot (e.g. sense of virtual community)—these concepts are represented as *latent* variables. A latent variable cannot be directly measured or observed, but instead is measured through a set of indicators or manifest variables. In our model, all constructs are "reflective" (as opposed to "formative"). Any change in a reflective construct is said to be "reflected" in its indicators (Hair *et al.*, 2019). That is, if the construct changes (which cannot be directly measured), it will 'cause' changes in its indicators, which are measured variables. Defining the constructs of studies such as ours

is particularly important given their latent (unobservable) nature, and links directly to the issue of construct validity (i.e., 'does the researcher measure what she intends to measure?'). A potential issue is that different studies may differently operationalize a given construct by defining different indicators. Further, particular care must be given to concerns of construct validity and discriminant validity so as to clearly define and distinguish related, but differing constructs.

We define the constructs of our model below, indicating what we mean by each construct, and through which indicators we measured them.

Sense of virtual community (SVC) was used as a latent variable that included the questions about dimensions of SVC.

Intrinsic motivations: We used the intrinsic motivations 1. Social motives (kinship and altruism) (Neel *et al.*, 2016); 2. Hedonic motives (joy and fun) (Tamilmani *et al.*, 2019); and 3. Moral motives (ideology and reciprocity) (Janoff-Bulman and Carnes, 2018).

English confidence was also used as a latent variable that included the four questions about self-confidence of fluency levels during interactions involving speaking and writing in technical and non-technical situations (Steinmacher *et al.*, 2021).

Power distance: We created an extra variable with the respective value of Power Distance index associated with each country, as proposed by (Hofstede, 2001)'s framework. The value can be accessed online [1] .

For the demographic questions, we leveraged and adapted from surveys used in OSS communities to ask about tenure, self-identified gender, and compensation (Bitergia, 2016; Corbet and Kroah-Hartman, 2017; Zlotnick, 2017b).

[1]https://www.hofstede-insights.com/fi/product/compare-countries/

Sample Analysis From the 316 received answers, we filtered the data to consider only valid responses for the theoretical model (i.e., those related to sense of virtual community, motivations, country of residence, gender, English confidence, and starting year at Linux Kernel). Respondents who did not complete the whole questionnaire were dropped (n=24). Next, we dropped the participants who answered "I'm not sure" to any of the items included with the five-point Likert scale for motivations (n=16) and sense of virtual community (n=51). After applying these filters, we were left with 225 valid responses, including residents of five different continents with a broad tenure distribution. The majority identified gender as men (84.4%), from Europe (52.9%), who were paid to contribute (65.4%), matching previously reported distributions of OSS contributors Zlotnick (2017b). Table 7.4 presents a summary of the demographics.

To establish an appropriate sample size, we conducted a power analysis, using the free G*Power tool Faul *et al.* (2009). We used an F-test with multiple linear regression, using an a priori test to compute the required sample size with a threshold value for medium effect size (0.25 Cohen (2013)), a significance level of 0.05, and a default value for the power (1 - β) of 0.95 Marcoulides and Saunders (2006). The maximum number of predictors in our model is six (three motivations and three control variables to SVC). This calculation indicated a minimum sample size of 62, which our sample of 225 considerably exceeded.

We used the software package SmartPLS, version 4, for the analyses. The analysis procedures for PLS-SEM comprise two main steps, each with tests and procedures. The first step is to evaluate the measurement model, which empirically assesses the relationships between the constructs and indicators (see Sec. 7.2.2). The second step is to evaluate the theoretical (or structural) model that represents the hypotheses (see Sec. 7.2.2).

Table 7.4: Demographics of the Linux Kernel respondents (n = 225)

Attribute	N	Percentage
Gender		
Man	190	84.4%
Woman	21	9.4%
Non-Binary	5	2.2%
Prefer not to say	8	3.6%
Prefer to self describe	1	0.4%
Continent of Residence		
Europe	119	52.9%
North America	68	30.2%
Asia	32	14.2%
South America	6	2.7%
Starting year at the Linux Kernel		
2000 or earlier	28	12.4%
Between 2001 and 2010	77	34.2%
Between 2011 and 2021	120	53.4%
Current Compensation for the Linux Kernel contributions		
Paid	145	64.4%
Unpaid (volunteer)	80	35.6%

7.2.2 Results

In this section, we describe our results, which include the evaluation of the measurement model (Sec. 7.2.2), followed by evaluation of the hypotheses in the structural model (Sec. 7.2.2), both computed through our survey data. We assess the significance of our model by following the evaluation protocol proposed by previous research Hair *et al.* (2019); Russo and Stol (2021) to make results consistent with our claims. The path weighting scheme was estimated using SmartPLS 4 (Sarstedt and Cheah, 2019).

Two tests are recommended to ensure that a dataset is suitable for factor analysis (Bartlett, 1950; Hair Jr *et al.*, 1995). We first conducted Bartlett's test of sphericity (Bartlett, 1950) on all constructs. We found a p-value $< .01$ (P values less than .05 indicate that factor analysis may be useful). Second, we calculated the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. Our result (.81) is well above the recommended threshold of .60 (Hair Jr *et al.*, 1995).

Evaluation of the Measurement Model Some of the constructs in the theoretical model (see Fig. 7.3) are modeled as latent variables, i.e., measured by more than one observed variable (i.e., item/question on the survey). The first step in evaluating a structural equation model is to assess the soundness of the measurement of these latent variables—this is referred to as evaluating the 'measurement model' (Hair *et al.*, 2019). We now present the assessment of several criteria.

Convergent Validity First, we assessed whether the questions (indicators) that represent each latent variable were understood by the respondents in the same way as they were intended by the designers of the questions (Kock, 2014), i.e., we assessed the convergent validity of the measurement instrument. The assessment of convergent validity relates to the degree to which a measure positively correlates with alternative measures of the same construct. Our model contains two latent variables, both of which are reflective (not formative), as functions of the latent construct. Changes in the theoretical, latent construct are reflected in changes in the indicator variables (Hair *et al.*, 2019).

We used two metrics to assess convergent validity: the average variance extracted (AVE) and the loading of an indicator onto its construct (the outer loading).

The AVE is equivalent to a construct's communality Hair *et al.* (2019), which is the

proportion of variance that is shared across indicators. A reflective construct is assumed to reflect (or "cause") any change in its indicators. The AVE should be at least .50, indicating that it explains most of the variation (i.e. 50% or more) in its indicators (Hair *et al.*, 2019). This variance is indicated by taking the squared value of an indicator's loading. All AVE values for both latent constructs in our model are above this threshold of .50.

A latent variable is measured by two or more indicators; indicators with loading below .4 should be removed because this implies that a change in the latent construct that it purportedly represents (or 'reflects') is not reflected in a sufficiently large change in the indicator (Hair *et al.*, 2019). Outer loading of .7 is widely considered sufficient, and .6 is considered sufficient for exploratory studies (Hair *et al.*, 2019). We followed an iterative process to evaluate the outer loading of the latent constructs; the indicators of the construct English confidence all exceeded .7, but SVC had two indicators below .7. We removed the SVC indicator, which had a loading below .4 (*svc6: a majority of the developers and I want the same thing*). After removing this indicator, the AVE value of SVC (now with five indicators) increased from .44 to .51 and all outer loadings were above .60.

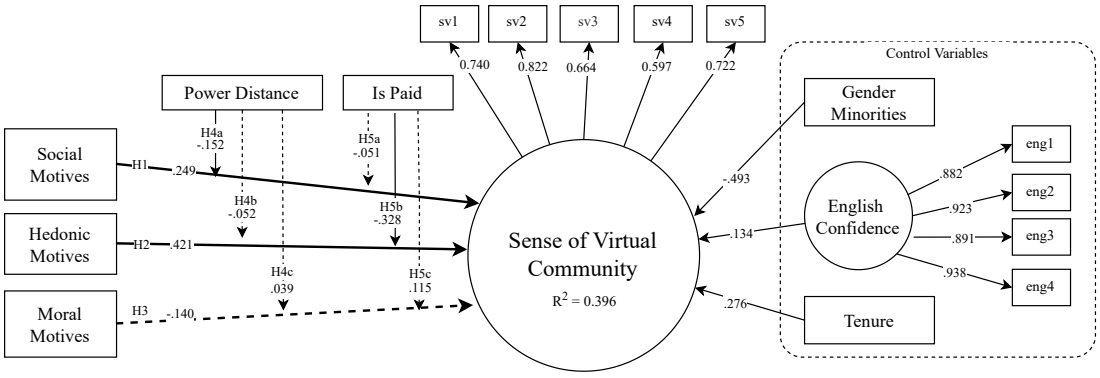

Figure 7.3: Item loadings and path coefficients ($p < 0.05$ indicated by a full line). Non-significant links are indicated with a dashed line

Internal Consistency Reliability Second, we verified how well the different indicators are consistent with one another and able to reliably and consistently measure the constructs, i.e., we assess the internal consistency reliability. A high degree of consistency means that the indicators refer to the same construct. There are several tests to measure internal consistency reliability. We performed both Cronbach's α and composite reliability (CR) tests; Cronbach's α frequently shows lower values, whereas CR is a more liberal test, which sometimes overestimates the values (Hair *et al.*, 2019).

For exploratory research, values of .6–.7 are acceptable for both Cronbach's α and CR, while for research in a more advanced stage values between .7 and .9 are recommended (Hair *et al.*, 2019). Values below .6 suggest a lack of internal consistency reliability, whereas values over .95 suggest that indicators are too similar and thus not desirable. Table 7.5 shows that both Cronbach's α and CR for the latent constructs fall within the range .75–.95, and their AVE is over .50.

Table 7.5: Internal Consistency Reliability

	Cronbach's Alpha	Composite Reliability	AVE
English Confidence	.927	.948	.820
Sense of Virtual Community	.761	.837	.510

Discriminant Validity Third, we verified whether each construct represents characteristics that are not measured by other constructs, i.e., we assessed the discriminant validity of the instrument (indicating the distinctiveness of the constructs). Our model includes two latent variables (SVC and English confidence). A primary means to assess discriminant validity is to investigate the heterotrait-monotrait (HTMT) ratio of correlations,

developed by Henseler *et al.* (2015). The discriminant validity could be considered problematic if the HTMT ratio exceeds .9 (Henseler *et al.*, 2015); some scholars recommend a more conservative cut-off of .85 (Hair *et al.*, 2019). The HTMT ratio between the two latent constructs (SVC and English confidence) was .24. We also assessed the cross-loadings of indicators and the Fornell-Larcker criterion.

Sense of virtual community (SVC) and English confidence are latent constructs, each measured with a set of indicators. We measured SVC through five indicators that covered the dimensions of feelings of membership, influence, belonging, mutual support, and emotional attachment. We measured English confidence through four indicators that covered performing reviews, speaking with others (face to face), and participating in technical and non-technical discussions on the email list.

Evaluation of the Theoretical Model We now evaluate and discuss the theoretical model, which involves the evaluation of the hypotheses.

Assessing Collinearity Our theoretical model has three different exogenous variables of intrinsic motivations, the moderators compensation and power distance, and the control variables English confidence, gender, and tenure. We hypothesized that the exogenous variables are associated with the endogenous variable sense of virtual community. To ensure that the three exogenous constructs are independent, we calculate their collinearity by means of the variance inflation factor (VIF). A widely accepted cut-off value for the VIF is 5 (Hair *et al.*, 2019), and in our model, all VIF values are below 5.

Path Coefficients and Significance PLS does not make any assumptions about the distribution (such as a normal distribution) of the data; therefore, any parametric tests of

significance should be used. To evaluate whether path coefficients are statistically significant, PLS packages employ a bootstrapping procedure. This involves drawing a large number (usually five thousand) of random subsamples with replacements. The replacement is needed to guarantee that all subsamples have the same number of observations as the original data set. The PLS path model is estimated for each subsample. From the resulting bootstrap distribution, a standard error can be determined (Hair *et al.*, 2019), which subsequently can be used to make statistical inferences. The mean path coefficient determined by bootstrapping can differ slightly from the path coefficient calculated directly from the sample; this variability is captured in the standard error of the sampling distribution of the mean.

Table 7.6 shows the results for our hypotheses, including the mean of the bootstrap distribution (*B*), the standard deviation (*SD*), the 95% confidence interval, and the p-values.

Based on these results, we found support for Hypotheses H1 (p=.002), H2 (p=.000), H4a (p=.045), and H5b (p=.023). Hypothesis H3 was not supported, nor were H4b, H4c, H5a, or H5b (all p values $> .2$). The three control variables all have significant associations with SVC: English confidence, gender, and tenure ($p < .05$).

Coefficient of Determination We assessed the relationship between constructs and the predictive capabilities of the model. The R^2 value of the endogenous variable in our model (SVC) was 0.4, which is considered weak-moderate (Hair *et al.*, 2019; Henseler *et al.*, 2009).

We also inspected Stone-Geisser's Q^2 (Stone, 1974) value, which is a measure of external validity, as an indicator of the model's predictive relevance (Hair *et al.*, 2019), and can be obtained through a so-called blindfolding procedure (available within the Smart-

Table 7.6: Standarized path coefficients, standard deviations, confidence intervals, and p values

	B	SD	95% CI	*p*
H1 Social motives→SVC	.249	.105	(.04, .46)	**.002**
H2 Hedonic motives→SVC	.421	.114	(.19, .64)	**.000**
H3 Moral motives→SVC	-.140	.112	(-.36, .08)	.215
H4a Power distance × social motives → SVC	-.152	.076	(-.31, -.01)	**.045**
H4b Power distance × hedonic motives →SVC	-.052	.074	(-.18, .11)	.477
H4c Power distance × moral motives→SVC	.042	.069	(-.10, .17)	.539
H5a is Paid × social motives→SVC	-.051	.065	(-.32, .20)	.696
H5b is Paid × hedonic motives→SVC	-.328	.144	(-.62, .05)	**.023**
H5c is Paid × moral motives→SVC	.115	.137	(-.17, .36)	.404
Gender minorities→SVC	-.493	.170	(-.81,-.14)	**.004**
English confidence→SVC	.134	.006	(.01,.25)	**.025**
Tenure→SVC	.276	.058	(.16,.38)	**.000**

PLS software). Blindfolding is a resampling technique that omits certain data, predicts the omitted data points, and then uses the prediction error to cross-validate the model estimates (Tenenhaus *et al.*, 2005). Q^2 values are calculated only for the SVC, the reflective endogenous construct of our model, with a value of .17. Values larger than 0 indicate the construct has predictive relevance, while negative values show the model does not perform better than would the simple average of the endogenous variable.

Moderating Factors We examined our data to determine if the impact of each intrinsic motivation on a sense of virtual community would change when they are exposed to a high power distance culture or when they are financially compensated to contribute.

Only results that were significant at 0.05 are reported, with confidence intervals calculated through bootstrapping.

- Power distance country culture: Being surrounded by a high power distance culture, in which leaders impose a high level of control and restrict the information flow Hofstede (2001), has been reported to negatively affect the sense of virtual community Ardichvili (2008). We did not find significant correlations between power distance and SVC for hedonic or moral motivations, but we found it for social motivations, which has a moderating effect on our model. Hence, we found support for H4a, but not H4b or H4c.

- Compensation: Being paid to contribute reduce the sense of virtual community for contributors driven by hedonic motivations, but not by social motivations nor moral motivations. Hence, we found support for H5b, but we reject H5a and H5c.

The three lines shown in Fig. 7.4 represent the slope for the relationship between social motives (x-axis) and SVC (y-axis) as moderated by power distance. This relationship is positive without the moderator (blue line going upwards from left to right). Before the interaction, the red line was below the blue line (mean). After the interaction, the red line was inverted to above the mean, meaning that the power distance weakens the positive effect of social motives on SVC, as the interaction occurs on the positive axis of both social motives and SVC.

Control Variables We also examined our data to determine if gender minority status, tenure, or English confidence strengthened or weakened the sense of virtual community. We found that participants who identify with gender minorities tend to have a lower sense of virtual community, while participants with higher tenure and English confidence reported a higher sense of virtual community.

Cluster Analysis: Detecting Unobserved Heterogeneity While moderators and context factors capture *observed* heterogeneity (see Sec. 7.2.2), there may also be *unobserved heterogeneity*, or *latent classes* of respondents, the presence of which could threaten the validity of results and conclusions Sarstedt *et al.* (2017). Latent classes of respondents refer to some groupings of respondents on one or more unmeasured criteria. The hypothesis results may differ for different groups.

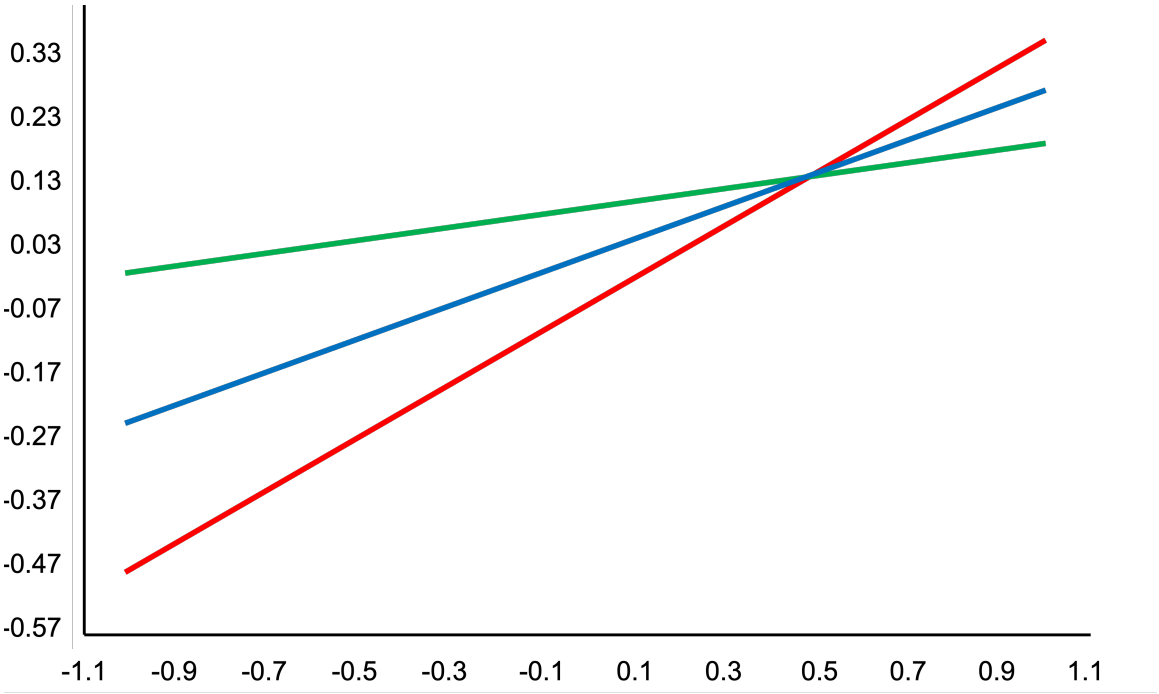

Figure 7.4: Power distance as a moderator of social motives → SVC

We adopted Becker et al.'s approach Becker *et al.* (2013), which jointly applies PLS-POS and FIMIX algorithms to identify latent classes. In Step 1, we used the minimum sample size for the maximum number of segments and ran FIMIX to find the optimal number of segments. In Step 2, we ran PLS-POS to compute the segmentation. In Step 3 we ran a multi-group analysis (PLS-MGA) and evaluated whether the segments were distinguishable. In Step 4 we checked if the resulting groups were plausible. We discuss the steps in more detail.

In Step 1, we assessed the maximum number of segments according to the minimum sample size (see Sec. 7.2.1). The minimum integer from dividing the sample size (225) by the minimum sample size (62) yields a theoretical upper bound of three; each segment should satisfy the minimum sample size. We ran FIMIX for one, two, and three segments (see Table 7.7) Sarstedt *et al.* (2017). The results were compared using several different retention criteria (see Table 7.7) Sarstedt *et al.* (2017). For each criterion, the optimal solution is the number of segments with the lowest value (in *italics* in Table 7.7), except in terms of EN, where higher values indicate a better separation of the segments. Sarstedt et al. Sarstedt *et al.* (2011) argue that researchers should start the fit analysis by jointly considering the combination of modified Akaike's information criterion with factor 3 (AIC3) and Consistent AIC (CAIC). More precisely, when both criteria suggest the same number of segments, this result is likely to be most appropriate. An alternative evaluation is whether modified Akaike's information criterion with factor 4 (AIC4) and Bayesian information criterion (BIC) suggest the same number of segments. In our case, neither AIC3+CAIC nor AIC4+BIC point to the same solution. The next option considers the joint analysis of Akaike's Information Criterion (AIC) and Minimum Description Length with factor 5 (MDL5); first, consider the number of segments indicated by the lowest values of AIC (3

Table 7.7: Establishing adequate number of segments

Combin	Criterion	1-Segment	**2-Segment**	3-Segment
1	AIC3	574.153	540.241	*517.508*
	CAIC	*625.395*	646.141	678.065
2	AIC4	589.153	571.241	*564.508*
	BIC	*610.395*	615.141	631.065
3	**AIC**	559.153	509.241	***470.508***
	MDL5	***935.361***	1286.737	1649.292
4	EN	0	0.869	0.821

segments) and MDL5 (1 segment). The appropriate number of segments should be lower than suggested by AIC (because it tends to overestimate), and higher than the number of segments suggested by MDL5 (because it tends to underestimate). Hence, this combination suggests that a 2-segment solution is appropriate because 2 is lower than the 3 suggested by AIC and higher than the 1 suggested by MLD5.

In Step 2, we evaluated the segment sizes of the 2-segment solution and proportions of data to check whether groups were substantial or candidates for exclusion. A segment is not substantial if its size is considerably lower in proportion (e.g., a 2% segment size) or below the minimum sample size Becker *et al.* (2013). The 2-segment solution divided the dataset into groups with 158 (70.2%) and 67 (29.8%) observations; both considerable portions and larger than the minimum sample size Becker *et al.* (2013).

In Step 3, we ran a multi-group analysis (PLS-MGA) with parametric tests to verify

whether the segments were distinguishable Becker *et al.* (2013), i.e., whether the results differ for the two segments. We found significant differences in hypotheses H4b-c, H5a-c, and on the control variables tenure and English confidence (see Table 7.8), thus we conclude these two segments represent two different groups of respondents. Both groups presented R^2, goodness-of-fit (GoF), and SRMR Sarstedt *et al.* (2017) equal or more favorable than the original model. The values of the path coefficients and the explained variance of the endogenous variable SVC are shown in Table 7.8, which presents the results for the two segments, as well as the original estimates (see column *B* in Table 7.6).

In Step 4, we examined that groups were "plausible" Becker *et al.* (2013) by explaining

Table 7.8: Group Paths Coefficients: coeficients in bold are significant; lines in gray show significant difference between segments

	2-segment solution		Orig.
	Seg. 1 Hedonic	Seg. 2 Social	All
Sample size (N)	158	67	225
Coefficient of determination (R^2)	.57	.94	.40
H1 Social motives → SVC	-.04	**.22**	**.25**
H2 Hedonic motives→SVC	**.31**	.06	**.42**
H3 Moral motives→SVC	-.03	**-.23**	-.14
H4a Power distance × social mot.→ SVC	-.10	**-.24**	**-.15**
H4b Power distance × hedonic mot.→SVC	**-.07**	.14	-.05
H4c Power distance × moral motives→SVC	-.02	.22	.04
H5a is Paid × social motives→SVC	**.49**	**-.61**	-.05
H5b is Paid × hedonic motives→SVC	**-.50**	**.50**	**-.33**
H5c is Paid × moral motives→SVC	**-.15**	.32	.12
Gender minorities→SVC	**-.70**	**-.92**	**-.49**
English confidence→SVC	-.15	**.88**	**.13**
Tenure→SVC	**.43**	-.05	**.28**

the different segments (highlighted in gray in Table 7.8) to label the segments. This labeling is somewhat speculative and by no means definitive, not dissimilar to the labeling of emergent factors in exploratory factor analysis. Given that for Segment 1 only Hedonic motives are significant, we posit that this segment represents *Hedonists* (*B*=.31); for Segment 2, we find that social motives are significant (*B*=.22), thus we label Segment 2 as *Socially Motivated*. We note that moral motives were not significant in the original analysis (see column 'Orig.'), but this did become significant with a negative coefficient (*B*=-0.23) for Segment 2. For the hedonists (Seg.1), tenure (*B*=.43) is positively associated with SVC. When social motives are associated with SVC (Seg. 2), English confidence positively affects SVC (*B*=.88). For both hedonists (*B*=-0.50) and socially motivated (*B*=-0.61) contributors, the association with SVC weakens when they are paid. Both groups showed that being a gender minority is associated with less SVC.

Evaluation of the Hypotheses We developed a theoretical model grounded in psychology literature for studying the relationship between a sense of virtual community and intrinsic motivations in OSS. The theoretical model includes a number of salient factors that have been shown to be important for belonging to an online community in general, but not yet within the OSS domain. Over the past two decades, the nature of OSS communities (as a specific type of online community) has changed; traditionally men-dominated and primarily consisting of volunteers, payment is now common, and increasingly we observe the participation of what we refer to as ''minorities' in the broadest sense of the word, including women. Our analysis highlights a number of key findings and implications; as we discuss these quantitative results, to illustrate the discussion we bring exemplar quotes from the respondents' responses to the final open question of the survey.

H1. Social motives → SVC: Social motives have a positive association with SVC. The intrinsically social motivations of kinship and altruism are positively associated with a sense of virtual community in OSS. This finding was corroborated by one of our respondents in the final open question, who associated SVC with social motivations: *"I did not fit in, in a big way. I was never able to create enough social capital to make networking effective, no matter who I tried to connect with."* Another respondent mentioned *"not being able to relate to colleagues* and named their perceived lack of SVC as *"a sense of otherness that never goes away."* However, the cluster analysis (Sec. 7.2.2) indicated non-significance for Segment 1 (which we labeled 'hedonic'), but significance for Segment 2 (labeled 'social'). We also found that for the 'socially motivated' English confidence is much more strongly related (B=.88 instead of .13) to SVC. This is intuitive because socially motivated people seek interaction, and English is the primary language within the Linux Kernel community.

H2. Hedonic Motives → SVC: Hedonic motives have a positive association with a sense of virtual community. OSS communities should seek to prevent toxic and other types of undesirable behavior that might reduce contributors' enjoyment; communities could also consider adopting clearer community codes of conductRaman *et al.* (2020); Cohen (2021); Miller *et al.* (2022). The cluster analysis showed that when only hedonism (not social motives) is associated with SVC (Seg. 1), tenure is also associated with SVC. Hedonic-motivated contributors from our sample also have longer tenure associated with SVC. Those contributors may have surpassed the initial barriers Steinmacher *et al.* (2014c) and find enjoyment, or, as mentioned by another respondent: *"It is therapeutic. When I feel bad about myself … it calms me down emotionally to do Kernel development."*

H3. Moral Motives → SVC: The cluster analysis did not support H3. While social motives are positively associated with SVC (Seg. 2), moral motives are negatively associ-

ated with and reduce SVC. The first association is expected, and not surprising Neel *et al.* (2016). People motivated by kinship or because they are happy to help others are keener to be part of the team and feel good in a community Kim and Drumwright (2016); Chang *et al.* (2016). Interestingly, the SVC presented a negative association with moral motivation. We argue that people motivated by ideological reasons may contribute regardless of how they feel about belonging. They do it because they feel it is the right thing to do, either because it is the most ethical choice, as advocated by the Free Software Foundation (https://www.fsf.org/), or because they have a moral debt to the software project that they use Janoff-Bulman and Carnes (2018). Future research can investigate how strong the ties between these people and the community are and what roles they play in SVC.

H4a/b. Power distance moderates the relationship between (a) social and (b) hedonic motives to SVC: Being surrounded by a country culture with a high level of power distance weakens SVC for socially motivated contributors (when we consider all contributors). Still, if we consider Seg. 1 (hedonic) on the cluster analysis, we observe that power distance also weakens the SVC associated with hedonism. These results align with cognitive theory Deci and Ryan (1985); contributors driven by hedonic (Seg. 1) or social motives (All) need more autonomy (through less hierarchy—less power distance) to develop a sense of virtual community. When not exerted in toxic and harsh ways to discipline community members, concerted control of communications can also ultimately play a prosocial role in increasing the SVC, by increasing cohesiveness, commitment, and conformity Gibbs *et al.* (2019).

H5a/b. Payment moderates the relationship between (a) social and (b) hedonic motives to SVC: Being paid to contribute weakens the association with SVC for hedonist contributors. Through the cluster analysis, we conclude that being paid to contribute also

weakens the SVC associated with social motives. Even though they enjoy contributing to the Linux Kernel, paid contributors driven by hedonic or social motivations showed a lower sense of virtual community in the Linux Kernel. This result aligns with cognitive theory Deci and Ryan (1985) and can be explained by the conflicting identities and hybrid belonging paid contributors have to both their sponsoring firms and the Linux Kernel community. We hypothesize that these contributors would leave the community if there was no payment to compensate for their participation.

7.2.3 Threats to Validity

We now discuss the threats to validity of this study.

Construct Validity. We adopted and tailored existing measurement instruments and developed derived measurement instruments for some constructs based on prior literature. Our analysis of the measurement model confirmed that our constructs were internally consistent, and scored satisfactorily on convergent and discriminant validity tests.

Internal Validity. Our hypotheses propose associations between different constructs rather than causal relationships, as the present study is a cross-sectional sample study Stol and Fitzgerald (2018). We acknowledge the limitation that our respondents comprise contributors who are more likely to have a sense of virtual community because they dedicated their time to answer the questionnaire, suggesting a response bias. While it is clear that contributors motivated by intrinsic-social reasons tend to experience a sense of virtual community, and that power distance and financial compensation can have an influence on those associations, a theoretical model such as ours cannot capture a complete and exhaustive list of factors. While other factors likely play a role, these results represent a useful starting point for future studies.

External Validity. We recognize the Linux Kernel is a mature project that attracts contributors for its value over the years, and so results may not be generalizable to all sizes and kinds of OSS project. We suggest further studies to replicate our findings. Our survey was conducted online and anonymously, but the numbers are aligned with the overall distribution of the Linux Kernel contributors. The Linux Kernel includes contributions from more than 15,000 developers, from over 1,500 companies Corbet and Kroah-Hartman (2017), and its contributors are mostly paid Corbet *et al.* (2012); Homscheid *et al.* (2015). According to previous research, around 10% of contributors of Linux Kernel identify themselves as women Bitergia (2016), and the majority of contributors are from the USA, which aligns with our sample. The responses were sufficiently consistent to find full or partial empirical support for four of our nine hypotheses.

7.2.4 Concluding Remarks

We found evidence that there is a subset of intrinsic motivations (social and hedonic motives) that are positively associated with the sense of virtual community (SVC); however, other extrinsic factors as the culture of the country and being paid to contribute can lessen SVC among contributors. Additionally, those who have higher English confidence feel a higher sense of belonging in the community, and contributors who identify as a gender minority tend to feel less of a sense of virtual community. Our results also show heterogeneity in our respondents, suggesting that there are different subgroups within the community for whom different motivations play a more prominent role. This suggests that a "one size fits all" approach would not work when designing interventions to create an inclusive, welcoming community.

In this chapter we investigated the attractiveness forces of motivations and sense of

virtual community. Next we will investigate the veering forces represented by the challenges faced by Linux Kernel contributors.

7.3 The Challenges Faced by Linux Kernel Contributors

The challenges faced by contributors exacerbate the already difficult task of contributing to OSS and can drain contributors' enthusiasm and motivation Sach *et al.* (2011). This difficulty can in turn affect the sustainability of the community.

Although sharing similarities with open collaboration communities, OSS communities have unique challenges. For example, the code base of an OSS project can be technically complex and require specific infrastructure to compile, run, and test. OSS projects also considerably differ from traditional software development organizations in terms of incentives, control, and coordination mechanisms Von Krogh *et al.* (2012). Traditional organizations have pay and career incentives, and other benefits stipulated as part of employment contracts Peters (2003). Finally, OSS contributions need a high degree of transparency in the form of visibility of actions on public artifacts and involve a community of geographically dispersed contributors Dabbish *et al.* (2012). The Linux Kernel, which is the case we studied, is a large ecosystem maintained by both volunteers and paid contributors who work for companies and foundations. Therefore, specific work is necessary to investigate the challenges to contribute to this unique environment.

Previous work has investigated the challenges in Apache, another large OSS project (Guizani *et al.*, 2021), OSS-specific contribution challenges (Steinmacher *et al.*, 2015b,a; Jensen *et al.*, 2011; Hannebauer and Gruhn, 2017), barriers faced by newcomers (Steinmacher *et al.*, 2015b, 2014a), one-time contributors Pinto *et al.* (2016), and mentors Balali *et al.* (2018).

In this study, we analyzed the challenges faced by Linux Kernel contributors, which we report in the present section.

7.3.1 Method

To conduct this study, we used the same data collected for the study presented in Chapter 7. Here we focused on a subset of questions that refer to the challenges faced in the contribution process. We detail the data curation and analysis for this study in the following.

Sample Analysis From the 316 answers received for the survey (see Section 7.1), we filtered the data to consider only the 197 responses that had valid responses for the challenges (163 men, 21 women, and 13 from other genders). The majority identified as men (82.7%), from Europe (71.1%), who were paid to contribute (59.9%), matching previously reported distributions of OSS contributors (Zlotnick, 2017b). The rate of non-blank answers (reporting at least one challenge) is almost the same for men and minorities: 62.2% (163 out of 262) of men and 61.8% of gender minorities (34 out of 55) reported at least one challenge.

Qualitative Analysis We analyzed the responses to the open question about the challenges faced while contributing to the Linux Kernel. The first author qualitatively analyzed the answers to the open questions by inductively applying open coding (Miles and Huberman, 1994) to organize what participants reported. We then organized our categories following the categories of challenges from a previous study (Balali *et al.*, 2018), include challenges related to Contribution Process (imposed by internal procedures or practices), Code-related or other technical challenges (related to or caused by tech-

nology, including frameworks, programming languages, and/or tools used to contribute), Interpersonal (related to the relationship among other contributors, maintainers, or community), and Personal (related to personal characteristics of contributors).

Segmented Analysis After completing the qualitative analysis, we checked the distribution of answers for each challenge. From the 197 respondents who reported they face some type(s) of challenges, 17 reported challenges related to more than one category, 20 reported only challenges related to the process, 50 reported only technical challenges, 79 reported only challenges related to interaction with other contributors (interpersonal), and 31 reported only personal challenges. We used descriptive statistics to summarize the responses and their association with the demographics data (Wohlin and Aurum, 2015).

To analyze how the challenges differ according to individual characteristics, we segmented our sample based on *tenure or years in the Linux Kernel*(fewer years in Kernel: $\leq$ 10 years in Kernel vs. more years in Kernel: $\geq$ 10 years in Kernel), *paid or unpaid*, and gender *men and minorities (women, non-binary, or prefer to self-describe.* We calculated the odds ratio for each challenge and demographic information. We interpreted the results as follows:

- if **Odds Ratio** $= \mathbf{1}$, both groups are equally distributed for the reported challenge.
- if **Odds Ratio** $> \mathbf{1}$, the likelihood for the reported challenge is higher for the first group (in our case: fewer years in Kernel, paid and men).
- if **Odds Ratio** $< \mathbf{1}$, the likelihood for the reported challenge is higher for the second group (in our case: more years in Kernel, unpaid, gender minorities).

7.3.2 *Challenges reported by the participants*

Our analysis of the responses revealed 14 challenges faced by Linux Kernel contributors. We organized these challenges using the four categories of challenges proposed by Balali *et al.* (2018) (process, interpersonal, technical, and personal), as can be seen in Figure 7.5. In the following, we present our findings organized by the categories of challenges.

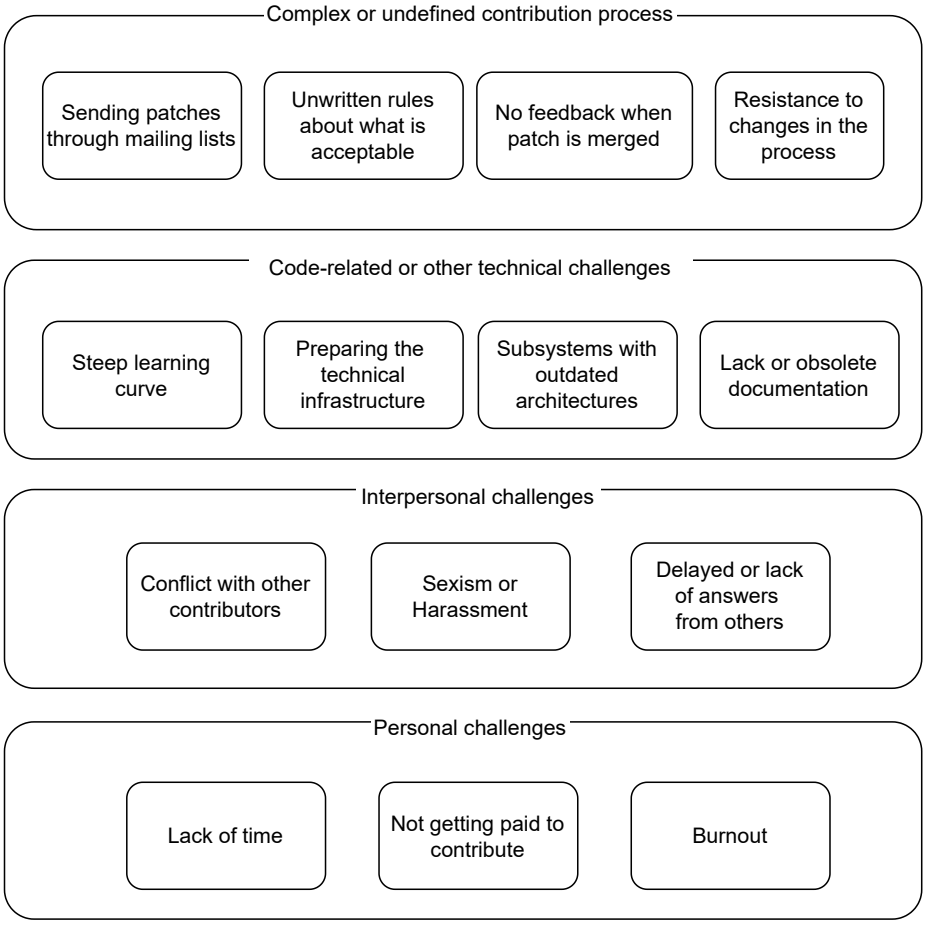

Figure 7.5: Challenges reported by Linux Kernel contributors

Contribution process challenges

Imposed by internal procedures or practices of the Linux Kernel Balali *et al.* (2018), the contribution process challenges emerge from a complex or undefined contribution process that is also hard to change.

Our participants mentioned there are UNWRITTEN RULES ABOUT WHAT IS ACCEPTABLE for patches when making contributions: *"Existing contributors' code gets easy adoption, while newcomers get quite a bit of dissection, which I feel is not uniform. There should be a uniform code of conduct for the code written irrespective of the author"* (S272). Regarding standards, *"maintainers or well-respected community members are effectively able to block unless code is designed and written the way they mandate"* (S109). Regarding size, while contributors often submit small contributions, maintainers find it overwhelming to handle a large number of small patches. Due to this lack of transparency about what is acceptable, contributors perceive their small contributions to be under-valued (S94) and that the contributions of developers who are well-known and trusted by maintainers are more easily accepted (S6, S92). The process of SENDING PATCHES THROUGH MAILING LISTS includes an email workflow and is considered a *"very difficult process for beginners, not very streamlined and daunting to get everything right even with several iterations"* (S47). This can be true not only for newcomers, but also for an experienced contributor, as S108 mentioned wasting time *"during three days figuring out what's wrong with the setup because it didn't work"*. Besides being complex, the process of sending patches through mailing lists also brings *"stress by having code being reviewed in a public mailing list"* (S58) and was mentioned as *"fragile and annoying"* (S18). Contributors also miss any kind of FEEDBACK THAT THE PATCH WAS MERGED as a final step of the contribution process (S215).

Our participants mentioned NOT FINDING SPACE TO PROPOSE CHANGES TO THE CON-

TRIBUTION PROCESS, as contributors face *"conservative reaction to new ideas"* (S25). Even maintainers face a hard time to *"convince others that the changes are good"* (S179), considering there are *"deeply entrenched maintainers/developers who resist any major changes"* (S262) and who *"want to avoid change rather than help[] find solutions"* (S132). Regarding changes to the software, S209 mentioned that "contributing to upstream drivers, improving the current code, fix bugs is fine. But providing new features or new framework[s[] is very hard."

Code-related or other technical challenges

Technical challenges are those related to or caused by technology, including frameworks, programming languages, and/or tools used to contribute to the Linux Kernel project (Balali *et al.*, 2018). This category of challenges included a steep learning curve, issues related to preparing the technical infrastructure to contribute, and a lack of or obsolete documentation.

"STEEP LEARNING CURVE" was a term repeatedly mentioned by our participants (S5, S67, S71, S112, S153, S201, S303). Beginners miss "good first issues" and a strict direction about where to start, as " there are so many places someone can start contributing, that it makes it difficult to decide where to begin" (S146). Besides the pathway to start, our participants mentioned the issues are too complex to solve and that *"subsystems are complicated and hard to quick-dive"* (S264). There is an ARDUOUS PATH TO FAMILIARIZE WITH ARCHITECTURE AND CODE BASE which include the architecture design and description of various subsystems (S99, S112). Contributors need also to familiarize themselves with the different ways to contribute to each subsystem (S63) and be able to understand a huge amount of code, in which different parts involve or influence each other (S112, S216).

Preparing the technical infrastructure to make a contribution was reported as involving too much effort to set up, install, and configure the environment for testing the code, which was *"a recurrent problem"* for S114, a newcomer. S178 reported that *"sometimes is hard to properly test things"* and S216 added that *"testing [in Linux Kernel] takes more setup effort than for other projects"*.

Subsystems with outdated architectures comprise another technical issue, as *"a lot of the subsystems [have] an x86 (or even Intel) bias that makes contributing other architectures a bit odd/harder"* (S214).

Technical hurdles worsen when there is also a lack of or obsolete documentation about the architecture and the code of the subsystems. The documentation is not organized, but *"scattered across commit messages of variable quality"* (S55). The obsolete documentation makes it *"hard to keep up with the rapid changes in the code base"* (S233). Newcomers usually suffer even more in finding resources for self-learning. Unwritten standards for each subsystem that vary per maintainer or *"according to what the maintainer likes"* (S159) necessitates having more contact with the maintainer who is responsible for the subsystem. Sometimes the code review brings an *"unwritten rule, as we no longer do that [type of code]"*. The frustration could be avoided with updated documentation. However, our participants reported that it was *"hard to find who is the responsible person for each subsystem* (S159) or *"the right contacts for unknown areas to upstream fixes"* (S195).

Interpersonal challenges

Challenges related to the relationship among other contributors, maintainers, or community were categorized as interpersonal (Balali *et al.*, 2018). They included sexism or

harassment, conflict with other contributors, and delayed or lack of answers from others, causing frustration.

Conflict with other contributors can represent a challenge when there are different goals between contributors and difficultly in reaching consensus (S152). Sometimes the conflict happens to *"arrive at a consensus on design topics"* (S272) or even in a more general way having a *"disagreement about what end users need"* (S171). Besides the conflict between different roles (maintainers and non-maintainers), our participants reported that it is hard to conciliate volunteers and paid contributors, who can have different interests and available time to dedicate to the contribution process (S159). Some maintainers are harsh during interactions, unable to balance between honest feedback and clashing, acting *"annoyed, abusive and breed[ing] a toxic environment with a bunch of walls setup to anyone contributing"* (S250). Contributors reported that maintainers cannot *"find a balance between code quality and nit-picking"* (S191). The toxic environment brings "fear of criticism" (S120), "shyness" (S101), and *"an embedded fear of public"* (S58), "as some members resort to public shaming to get the contributor to not make mistakes and all this does is breed fear" (S250). Conflicts have consequences; participants mentioned developing a *"social phobia* (S64) that can also hinder contributions, because *"sending patches was always stressful ... so [it was preferable to] contribut[e] only patches that were obviously correct and didn't need any discussion"* (S64).

Sexism or harassment were reported by both S37 and S236 as *"frenemy issues among women in tech"* (S37) and *"harassment based on gender* (S177, S226).

Delayed or lack of answers from others during code reviews causes frustration. Answering questions with non-actionable and highly demanding requests during code reviews is often time-consuming, such as *"pushing logic deeper into the core of the ker-*

nel, where the bar for inclusion is too high for acceptance" (S132). Besides the hostility and toxic environment (S223, S241, S250, S268, S288, S308), contributors mentioned that emails with questions and code reviews experience *"long delays* (S76, S152, S222, S254) before being answered. The delay can be longer when the contributor who submitted the patch is not known to the reviewer (S92). Some participants explained the delays of maintainers' answers due to limited reviewer resources (S76, S100), and then *"when the top-level maintainer is out/busy, the whole process is stuck"* (S12).

Personal challenges

We categorized personal challenges as those related to personal characteristics of contributors (Balali *et al.*, 2018), including lack of available time to make contributions, not receiving monetary incentives to make contributions, and feelings of burnout.

Lack of time was reported by both paid and unpaid contributors. Paid contributors reported a challenge to simultaneously contribute to the Kernel while working on other tasks (S187, S190, S194, S206, S308). Volunteers also use their spare time to study other technologies (S33).

Not getting paid to contribute represents a challenge that can hinder contributions from both maintainers and non-maintainers. While maintainers are *"often not funded"* (S96), non-maintainers face difficulties to be hired to make a living from OSS, even if they are experienced (S17).

Burnout was a feeling mentioned by both maintainers and non-maintainers. Maintainers consider that their efforts are usually not recognized. Both maintainers and non-maintainers struggle with burnout due to hostility. Maintainers face *"anxiety about being yelled at by Linus for some random minor detail"* (S180). Non-maintainers reported

burnout due to long delays for reviews, described by S86 in terms of: *"Feeling obligation-driven, meaning they end up with a steadily growing workload. It can be very dispiriting for people to work hard on something for a long time and then either get negative feedback or nothing whatsoever".*

A segmented look at the challenges perceived by the Linux Kernel contributors

In addition to the categorization described above, we took a deeper look into the results to understand the prevailing reports of challenges among our respondents and across different demographics. We avoid using the numerical prevalence of evidence to indicate the importance or criticality of any challenge. However, when presenting the results, we use supplementary and corroborative counting of the responses to triangulate the qualitative analysis Hannah and Lautsch (2011). The majority of respondents reported interpersonal challenges (49.7%), followed by technical (35.2%), personal (19.6%), and process (13.7%). Figure 7.6 illustrates the categories of challenges for each demographic.

The percentages in the Figure 7.6 reflect the number of participants who mentioned any challenge under each category of challenge. Respondents often provided challenges from multiple categories.

We also examined the intersections of demographics. Of the 34 respondents in gender minorities, 61.8% (21) are paid, 38.2% (13) have a tenure of more than 10 years in the Linux Kernel, and 17.6% (6) are both currently paid and have contributed to the Linux Kernel for more than 10 years. Those paid and experienced contributors reported only interpersonal and personal challenges.

From the 163 men, 59.5% (97) are paid, 49.7% have a tenure of more than 10 years in the Linux Kernel (81), and 27.0% (44) are both currently paid and have contributed to the

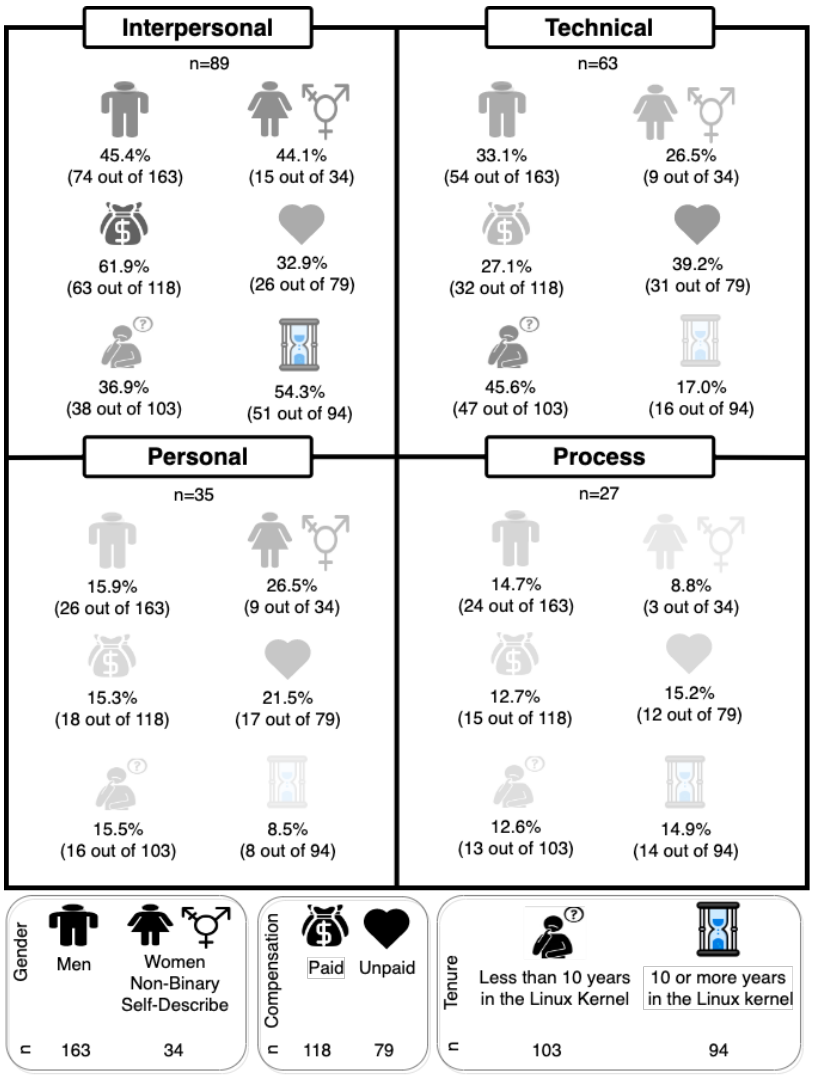

Figure 7.6: Subgroup analysis of challenges. The opacity of the icons represents the percentage of each group for the category. Darker means a higher percentage. Some respondents provided answers about challenges in more than one category.

Linux Kernel for more than 10 years. Different from the gender minorities, the group of men who are also paid and experienced still reported the four categories of challenges.

We also calculated the odds ratio for each of the four categories of challenges, considering (i) gender (men or gender minority groups); (ii) compensation (being paid or unpaid to contribute); and (iii) experience in the Linux kernel (less or more than 10 years).

Table 7.9 presents the results of the odds ratio for each category of challenge. According to our sample, paid contributors have higher odds (2.3x) than unpaid contributors to report interpersonal challenges. Moreover, unpaid contributors have greater odds (1.7x) to report technical challenges than those who are paid. We also found that more experienced contributors have greater odds (2.04x) than less experienced contributors, reporting both interpersonal and technical challenges.

Table 7.9: Odds ratios of challenges per personal characteristic

	Men vs. Gender Minorities	**Less vs. More Experienced**	**Paid vs. Unpaid**
Interpersonal Challenges	1.05	0.49**	2.33***
Technical Challenges	1.38	0.49**	0.58*
Personal Challenges	0.72	0.73	0.53
Process Challenges	1.78	0.83	0.81

Significance codes: * $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$.

Note: Odds ratio > 1 means that the first segment has greater chances of reporting the challenge than the second. Ratio < 1 means the opposite. The challenges were coded from the open question.

7.3.3 Strategies suggested to mitigate the challenges

Even though we did not ask about how to mitigate the challenges, some respondents provided ideas and mechanisms related to the challenges they have faced. We present these recommendations here, which would be useful for guiding future efforts in the Linux Kernel community.

Mentorship Both S4 and S30 mentioned facing technical challenges, but having "mentors helped to make this challenge manageable" (S30). S4 added that she *"was fortunate to have mentors who shared valuable resources like books, talks, and blogs that kept me going.* S271 lacked a mentor, but desired one, noting that mentorship *"is especially important in*

the (many) areas of the kernel whose internals are poorly documented" .

Training [2])is an online programming exercise on how to get a patch accepted in the Linux kernel, which starts from very basic modules and increases in complexity. This training was mentioned by S80 as helpful to navigate and mitigate the challenges of the contribution process.

Change the contribution process by "integrating email/cmdline" was mentioned by S103 as necessary to facilitate the contribution process and avoid "people get[ting] discouraged before they even start".

More objectivity in the feedback was suggested by S288, as she reported receiving answers that did not make clear what should be adjusted. Moreover, S215 asked for "a short text like 'patch accepted' for merged patches."

7.4 Threats to Validity

There are some limitations related to our research results.

Survival bias. Our results reflect the opinion of current contributors. Therefore, to increase inclusiveness by fully understanding the reasons contributors might leave, we acknowledge that additional research is necessary to understand the point of view of the contributors who left the Linux Kernel.

Recall bias. As our question was intentionally open-ended, our results could be impacted by either salience bias, where respondents focus on definitions that are prominent or emotionally striking and not necessarily all the factors that matter; or by memory bias, where participants answered questions based on what they can first recall. However, topics that are relevant to the respondent often emerge from the spontaneous answers.

[2] http://eudyptula-challenge.org/

Data Consistency. Consistency refers to ensuring that the results consistently follow from the data and there is no inference that cannot be supported after data analysis Merriam and Tisdell (2015). The group of researchers performed the qualitative analysis of questionnaire responses. We had weekly meetings to discuss and adjust codes and categories until reaching agreement. In the meetings, we also checked the consistency of our interpretations. All analysis was thoroughly grounded in the data collected and exhaustively discussed among the whole team. The team includes researchers with extensive experience in qualitative methods.

Theoretical saturation. A potential limitation in qualitative studies regards reaching theoretical saturation. From contributors in this study with different perceptions about the studied phenomenon, we received 197 responses for the challenges question. The participants were diverse in terms of experience, country, gender, and financial compensation. Therefore, although theoretical saturation cannot be claimed, we believe that we obtained a consistent and comprehensive account of the phenomenon for the studied case. After analyzing the 50^{th} answer to the challenges question we did not find any new categories, and proceeded using the existing categories for the following 147 challenges.

7.5 Concluding Remarks

The majority of Linux Kernel contributors reported interpersonal challenges (49.7%), followed by code-related or other technical challenges (35.2%), personal (19.6%), and contribution process (13.7%). Paid contributors have higher odds than those who are unpaid to report interpersonal challenges. The opposite happens for technical challenges, which unpaid contributors have greater odds to report than those who are paid. More experienced contributors have greater odds than less experienced contributors to report both

interpersonal and technical challenges.

In contrast to what we expected regarding the gender perspective, we found no difference in the odds when we compared men and minorities. Interestingly, when we considered only experienced contributors (with more than 10 years in Linux) who are paid, we observed that people from gender minorities do not report any technical or process-related challenges. This suggests that, even while consistently reporting a lack of self-confidence, tenure gives them the necessary knowledge to overcome technical and process barriers.

Chapter 8

DISCUSSION

In this chapter, we discuss the findings of this dissertation. The goal is to tie together the perspectives brought by every study in a single place, providing a more complete picture of the outcomes of the dissertation.

8.1 Considering women contributor's motivations, perceptions of being successful, and challenges

8.1.1 Motivation to join vs. project culture

Since women have social motivations (e.g., Kinship) as we presented in Section 3.2, and the reported challenges are also social, as we presented in Section 3.4, there is a conflict between their expectations and reality, which can explain why women are not joining or staying in OSS projects. When women join an OSS project expecting to find other women (Paul *et al.*, 2019) and friendly colleagues (Prana *et al.*, 2021), but instead find Lack of Peer Parity and face a Toxic Culture, this conflicts with their motivations. Enjoyment was not found as one of the top motivators for women, which could be because of the lack of psychological safety in the Toxic Culture. So women may not feel safe having fun while contributing, and fun is not their top motivation. One place toxicity can manifest in communications is via comments in code reviews and mailing lists. Code reviewers may need education support to articulate their review comments in a way that builds relationships (Bosu *et al.*, 2016). One strategy could be providing *review templates* that support developers to use inclusive words and employ empathy.

The lack of peer parity can be alleviated by attracting more women, which can be accomplished by recognizing different types of contributions (Trinkenreich *et al.*, 2020b), for example, recognizing contributors who participate by answering questions and discussing issues (Trinkenreich *et al.*, 2021b; Ducheneaut, 2005). Moreover, communities can foster peer communication through women-(and ally-)only groups and events, such as R-Ladies and other safe spaces (Canedo *et al.*, 2020; Singh, 2019b,a). The discrepancy between aiming to collaborate and encountering a Toxic Culture that leads to problems of impostor syndrome (Section 3.4) can push women away from an OSS project. This is an example of how different factors (i.e., motivations and challenges) may be considered together to understand women's participation in OSS.

8.1.2 Future goals vs. benefits received from contributing

Considering our participants, recognition was more mentioned by women than by men as a career goal when we asked about their perceptions of being successful 6. The discrepancy between aiming to be valued and not being recognized when playing non-coding roles 4 can also lead to problems of impostor syndrome (Section 3.4) and represent an additional factor that pushes women away from an OSS project. While coders gain recognition from having their names in a "credits" file or badges in their profiles, non-coders are commonly overlooked because their activities are harder to quantify, as they bring benefits that many times are intangible or by their nature difficult to measure Trinkenreich *et al.* (2021a). This is another example of how different factors (in this, perceptions of being successful and challenges) may be considered together to understand women's participation in OSS.

8.1.3 OSS as a career pathway

As Section 3.2 presents, only 4.07% of the 226 surveyed women from a FLOSS 2013 study joined to increase their job opportunities. After becoming contributors, this motivation increased almost six times (to 25.79%) (Robles *et al.*, 2016). We argue that this represents the "shifting belief" that women have in OSS toward building a career, which increases only after overcoming the barriers to joining and becoming contributors. The multiple roles presented in Section 3.3 relate both to the technical (project-centric) and non-technical (community-centric) sides of the projects. Therefore, an awareness of the different roles and career pathways in OSS can attract women with diverse backgrounds and expertise to OSS by showing them the multitude of trajectories to success Trinkenreich *et al.* (2020b). Programs like Google Summer of Code, as well as other OSS-academic liaisons, can improve awareness of career opportunities provided by OSS. Further, given that 54% of the women who contribute to OSS devote less than 5 hours per week, the majority do not make a living from OSS. The fact that WORK-LIFE BALANCE ISSUES are a challenge that women face (see Section 3.4) and BEING PAID TO CONTRIBUTE is a relevant motivation (see Section 3.2), OSS projects could offer part-time jobs to attract women who are not yet participating and ensure that women are in positions to mentor other women.

8.2 Implications for Practice

8.2.1 Combining synergistic strategies to mitigate the challenges faced by women

One option for communities looking to improve diversity is to combine synergistic strategies. This might be especially useful since the literature has identified many different strategies, which have low strength of evidence. OSS communities can start by im-

plementing simple, but structured actions combining ideas from more than one strategy. For example, by publishing success stories of women in the media, OSS communities can PROMOTE AWARENESS OF PRESENCE OF PEERS to attract more women and also RECOGNIZE WOMEN'S ACHIEVEMENT (VISIBILITY) to retain women who are already contributors. Considering that this media exposure can include women's posts and pictures, this action also helps with the strategy of DE-STEREOTYPE THE OSS CONTRIBUTOR, which has been associated with images of men in technical textbooks (Makarova and Herzog, 2015; Lee, 2018) and search results (Kay *et al.*, 2015). Another action that can use more than one strategy is to create a women-only forum, which is part of the strategy to PROMOTE WOMEN-SPECIFIC GROUPS AND EVENTS. When moderating and analyzing the messages from women to implement feasible changes to problems that are being actively discussed, this action also acts to ENCOURAGE AND BE WELCOMING TO WOMEN by offering mentorship or inviting women to contribute to specific activities. Another action that can be adopted toward multiple strategies is to CREATE AND ENFORCE A CODE OF CONDUCT by providing online training on enforcement and being transparent about the punishments for those who violate the code of conduct. There can be a training for contributors in general, another for mentors to (PREPARE MENTORS TO GUIDE WOMEN), and a third for allies to advocate for women and act as "collaborators, accomplices, and co-conspirators" (Melaku *et al.*, 2020). The content of the training can include practical examples of acceptable and non-acceptable behaviors. Communities can use mining tools to identify gender pronouns in messages of mailing lists, pull requests, and code reviews and help to PROMOTE INCLUSIVE LANGUAGE.

8.2.2 Encouraging women to join OSS

Women newcomers can become aware of the different types of contributions made by other women through our study. Although not restricted to the activities presented here, women can be inspired by other women's success and motivations to participate, gain awareness of the challenges reported by other women so they can be prepared to face similar ones, and prioritize participation on projects that follow one or more reported strategies. Members who participate in a virtual community by exchanging information and providing support tend to develop more positive feelings toward the community and a stronger attachment to it Blanchard and Markus (2004). Community members can encourage newcomers to become more active and move beyond the stage of 'lurker,' enticing them to participate in mailing lists Tonteri *et al.* (2011) and start making social connections to establish mutual trust, be known by other contributors, and facilitate the development of their sense of virtual community. Conferences and meetups can help contributors who are hedonic and socially motivated have fun and increase their social capital.

8.2.3 Cultivating a sense of virtual community

SVC can be developed through *exchanging support* Blanchard *et al.* (2011); Tonteri *et al.* (2011), *creating identities and making identifications* Blanchard *et al.* (2011), *producing mutual cognitive and affective trust* amongst members of a community Blanchard *et al.* (2011); Chih *et al.* (2017), and establishing norms and a "concertive (but not enforced) control" Gibbs *et al.* (2019), in which members of the community become responsible for directing their work and monitoring themselves. Besides online interest groups for members, chat rooms, instant messaging, and discussion forums to encourage community involvement

Xu and Li (2015), OSS communities can provide online tools with shared spaces for contributors to work "together" on issues, and to discuss and collaborate on similar interests. Better interactions can strengthen contributors' sense of virtual community, especially for those who seek social relationships. When the information being exchanged surpasses the technical content and includes also socio-emotional support, it can evidence personal relationships among group members, and finally bring feelings of acceptance by members Blanchard *et al.* (2011). OSS communities should foster peer support among members to bring a positive impact on developing SVC Tonteri *et al.* (2011). Peer support includes both technical and social support and happens through comments in pull requests and participation in mailing lists (by either reading or posting messages). Communities can manage pull requests and mailing lists to guarantee that members' posts are not being missed Miller *et al.* (2022), and that the communication adheres to the code of conduct.

Opportunities for education: Once they are aware of the challenges that women face, educators can address the underlying issues causing these challenges in the classroom, thereby improving students' (all gender) awareness of biases and discuss possible mitigation actions. This research can also inform educators who adopt contributions to OSS projects as a method to teach software engineering Pinto *et al.* (2017).

8.3 Limitations

8.3.1 External Validity

Generalizing results in OSS is not an easy task, given the diversity of ecosystems, governance models, languages, and composition of the projects. We recognize that our work on the Linux Kernel (Chapter 7) may not be generalizable to any size and kind of

OSS project, given that the Linux Kernel is a mature project that attracts contributors for its value over the years, and have specific governance model and processes.

However, we attempted to reduce this limitation in the other studies that are part of the present dissertation by inviting participants in the large (Chapters 4, 5, and 6). The interviews for Career Pathways (Chapters 4) and Perceptions of Being Successful (Chapter 6) studies comprised paid and volunteer contributors across different OSS projects (e.g., Kubernetes, Drupal, R, Noosfero, SPDX, envoy) that varied on size from 30 to 3,000 contributors, had different domains—including infrastructure and user-application projects, and types—backed by foundations, communities, and companies (e.g., Microsoft, Linux Foundation, Google, Red Hat, IBM, Drupal, Bitergia, Apache). The recruitment for surveys used as data collection for the Motivations (Chapter 5) and Perceptions of Being Successful (Chapter **??** studies involved researchers (who live in different countries) sending direct messages to their contacts and posting ads on social network websites, and advertising in social media. On Twitter, our posts were retweeted more than 200 times (Gerosa *et al.*, 2021). The survey respondents included both paid and volunteers respondents who contribute to Linux kernel, KDE, Debian, Kubernetes, LibreOffice, Mozilla, PHP, Laravel, Drupal, Debian, TensorFlow, Apache projects, Firefox, Homebrew, Arduino, Eclipse, Joomla, Django, WordPress, JavaScript libraries, Python libraries, and R packages. The projects are diverse in programming languages, age, community size, organization, and governance model (Gerosa *et al.*, 2021; Trinkenreich *et al.*, 2021a).

8.3.2 Data Consistency

Consistency refers to ensuring that the results consistently follow from the data and there is no inference that cannot be supported after the data analysis (Merriam and Tis-

dell, 2015). Regarding positionality, as we stated in Section 1.5, the researcher is partially an insider as she identifies the gender as a woman and has previously worked in the software industry. While the insider positionality offered the researcher benefits of empathy with other women and some facility to establish rapport and trustworthiness with interviewees, it also can bias the qualitative analysis. All analysis was thoroughly grounded in the data collected and exhaustively discussed amongst the whole team, including researchers who do not identify as women to balance this bias. The group used card-sorting techniques during open coding to reach a consensus during the qualitative analysis of studies presented in Chapters 4, 5, 6, and 7.3.

8.3.3 Survival Bias

The characteristics of our sample may have influenced our results. Our results reflect the opinion of current contributors. So, there may be reasons to leave the project, and reasons that could entice future contributors from those who are not currently contributing, either because they left or never joined. We acknowledge that these perceptions are not completely covered in this dissertation. To reduce the limitation related to the reasons to leave, in the survey (Chapter 7 we asked people about what would make them leave.

Chapter 9

CONCLUSION

Through the development of this work we collected evidence about women's participation in OSS and examined different factors that influence their involvement. The rates of women in OSS are increasing over time, but at a slow pace. Many organizations now have business goals to increase the rates of women in their teams and in leadership positions. Women desire opportunities, but not those rooted in benevolent sexism—i.e., only because they are women—that often brings sexist antipathy. They want equal conditions, treatment, and opportunities for realizing their full potential.

The study about career pathways (Chapter 4) unveiled community-centric roles, which are often hidden and unrecognized, and, according to previous studies, are currently more often held by women (3.3). Women have been demonstrating their potential as coders, but also as non-coders: as advocates, strategists, community managers, community founders, mentors, license managers, writers, and treasurers. While these roles do not produce code, they are important for the growth and sustenance of OSS, especially in the new OSS landscape.

The study about motivations (Chapter 4) evidenced that social motives, such as helping others and teamwork, are more relevant now than 20 years ago. Moreover, the study showed the complex and dynamic nature of motivations, as contributors usually have more than one motivation and often change their motivations over time, joining for one set of reasons and staying for another. OSS contributors often join due to extrinsic factors, and continue because of intrinsic factors. We noticed that men's odds to contribute due

to hedonic motives are 4x higher than women's, which corroborates a previous study from Burnett *et al.* (2010) that concludes women are generally more motivated to use technology to accomplish a goal rather than for fun. We discussed in Section 8.1 that the toxic culture faced by women may cause them to lack the psychological safety in OSS to contribute for enjoyment and fun in the same ways men do.

Following the concept of shifting motivations, we denoted the perceptions of being a successful OSS contributor (Chapter 6). The multifaceted definition of success included both objective metrics and subjective perceptions of accomplishments. We found that the interpersonal dimension plays a dominant role in the definition of success of contributors across different genders, in which factual and perceived contributions are the most referenced, followed by recognition. The definition of success can be used in future studies to investigate how it represents the career aspirations of contributors of different genders by driving their decisions to stay or leave an OSS project.

Belonging to an OSS community was reported as one of the intrinsic motivations to contribute (kinship), a perception of being successful (to bring satisfaction), and a challenge (when not "fitting with" a community). We investigated feelings of belonging as an additional factor that influences participation through the sense of virtual community. We showed the positive association between intrinsic motivations (social and hedonic motives), tenure, and English confidence and the sense of virtual community, and revealed that contributors from gender minorities (including women and non-binary people) tend to feel less sense of virtual community. The heterogeneity in our respondents suggested that the "one size fits all" approach would not work when designing interventions to create an inclusive, welcoming community.

It is clear that there is still a large gender disparity among OSS contributors. However,

the gender disparity is less pronounced in the initial stages (e.g., as students of Google Summer of Code). There is attrition of women contributors as they move through their pathways, making them underrepresented in core and leadership roles (Section 3.6.1). This attrition can be a consequence of the several socio-cultural challenges faced by women during the process. While mentorship events enhance (women) participants' sense of competence and increase the chances of future contributions' values, these programs alone are insufficient as women do not stay long enough to become project leaders.

The majority of the challenges that women face or the reasons that women leave OSS are socio-cultural in nature and unrelated to technical skills. Recognition and empowerment are strong strategies to fight the impostor syndrome, but while attaining promotion to leadership is not easy, serving in a high position can be even more challenging. Women often lack the support or authority to accomplish their strategic goals. Empowering women is not only about promoting them to leadership but preparing them for the position and also giving them the proper authority. The lack of peer parity can be alleviated by attracting more women, which can be accomplished by understanding they may have different motivations and career goals to reach their pots of gold, as well as recognizing different types of contributions Trinkenreich *et al.* (2020a). We observed a recent growth in the creation of a code of conduct in OSS projects. But it is unclear how these are enforced, and, for them to work, enforcement is key. While having a code of conduct will not prevent sexism, it indicates to everyone who demonstrates sexist behaviors that such actions will not be tolerated in the project. Communities should put mechanisms in place to implement the code and show that violations have consequences.

There is still a long work ahead for OSS, for the software industry, and for us as a society to create more diverse and inclusive environments. We hope to enlighten actions

towards reducing the perceived challenges and (more importantly) increasing awareness about the structural and cultural hurdles imposed on women that negatively influence gender diversity in OSS.

9.1 Future Work

This subsection discusses the gaps in the literature that may be explored in future research.

Intersectionality considers the simultaneous and mutually constitutive effects of the multiple social categories of identity, difference, and disadvantage (Cole, 2009), paying close attention to racism, sexism, and classism as they operate simultaneously (Haynes *et al.*, 2020). Future work can work to understand the intersectionalities of contributors to shift the focus away from individual-level conceptualizations of gender in OSS and toward structural examinations that take into account the power dimensions of race, class, culture, sexuality, caregiving responsibilities, disabilities, and other demographics, and how different systems of oppression are mutually constituted and work together to influence OSS contributors' participation. Sensitivity to intersections enhances insight into the issues of inequality, thus maximizing the chance of social change (Atewologun, 2018). Although being a convergence of factors, intersectionality research is not only about adding a given number of demographic variables (MacKinnon, 2013). Empirically investigating intersectionality is challenging. First, there can be a mismatch between identities (how women see themselves) and the demographic survey questions used. How women see themselves and with which social aspects women identify is something that should be investigated before creating the survey (Hughes and Dubrow, 2018). Secondly, the perennial "small-N" problem and unanswered questions can harm the ability to make statistical in-

ferences, as there can be too few observations in the sample to permit the desired analysis (Bauer, 2014). Finally, the researcher needs to find a suitable statistical technique to analyze the data when having relevant observations. The most basic and common approach to intersectional research is testing statistical interactions between two or more variables. However, more sophisticated approaches are being used in social and health research, such as the use of hierarchical linear models (HLM) or multilevel models (Fehrenbacher and Patel, 2020). HLM provides many advantages over ordinary least squares regression, as it helps to address complexity and variation within groups and between levels of analysis (e.g., for women of color with caregiving responsibilities who live in poor neighborhoods) (Evans *et al.*, 2018).

Evaluation of recommended strategies to increase women's participation: Although several strategies to increase women's participation have been proposed in the literature, few works present scientific evidence about their effectiveness. For instance, (Tourani *et al.*, 2017; Imtiaz *et al.*, 2019; Singh and Brandon, 2019) relegate the evaluation of effectiveness of the "code of conduct" to future research, despite the fact that it is one of the most-cited strategies to promote women's participation. Izquierdo *et al.* (2018) discuss the difficulty of evaluating the effectiveness of strategies, as communities need to have consistent measurements before (baseline), during, and after their implementation. The authors reported that although OpenStack created the Women of OpenStack Working Group (which included educational sessions, professional networking, mentorship, social inclusion, and enhanced resource access), the OpenStack Foundation lacked baseline information about the involvement of women.

The interplay between perceptions of being successful and motivations to contribute. Perceptions of success represent long-term goals and an imagined future career Frank (1938); Lewin (1936), which influences commitment Visagie and Koekemoer (2014) and human behavior Frank (1938); Lewin (1936). Future work can investigate how perceptions of success influence the motivations and retention of OSS contributors.

Recognition of community-centric roles is important to support the growth of people whose background is not related to software development. Their activities are harder to quantify, given that they usually do not leave traces on project repositories. This may pose challenges beyond proposing metrics and toward proposing changes in terms of how these activities are performed, logged, and weighted. Future work can include mechanisms to identify and measure the evolution of stakeholders performing non-technical roles, toward raising awareness and recognizing the non-code contributions.

Perspectives of possible future contributors We examined the factors that influence the participation of current contributors in OSS projects. While the perspective of contributors who are inside the ecosystem is important, future work can include the perspectives of those who have not yet joined OSS to understand what would entice them to become an OSS contributor.

Theories to explain why women leave or avoid OSS projects: The literature also reports a diverse set of challenges faced by women 3.4, but few make a theoretical connection as to why women leave (or avoid) OSS projects. Theoretical understandings can help create more effective, longer-term solutions. Some studies have analyzed motivation to participate in OSS projects (Lakhani and Wolf, 2003; Ghosh *et al.*, 2002; Hars and Ou,

2004; Hertel *et al.*, 2003), but only a few report women's motivation, and none go deep in this analysis. Moreover, the literature lacks research exploring why women leave OSS, their motivation to avoid participating in OSS, and why a large portion of women who study STEM does not join OSS projects. One of the few studies that have used theory to explain these phenomena is Qiu *et al.* (2019b), who found that establishing networks of relationships can support the long-term engagement of both men and women in OSS projects, and that when team members have more diverse programming language backgrounds, women are less likely to leave the project early.

Evaluation and expansion of the theoretical model in other communities We developed a theoretical model of the antecedents of the sense of virtual community that included intrinsic motivations. Future work could analyze other antecedents that can be positively or negatively associated with SVC, like challenges and perceptions of success. The control variables can include more demographics, such as types of contributions, race, and ethnicity. Moreover, while we investigated antecedents for SVC, future work can increase the path analysis of the theoretical model by investigating the consequences of SVC, such as intentions to leave and levels of contributions. While we created the SVC theoretical model based on a survey in the Linux Kernel, future work could replicate the survey to compare results in another community. We are working with the Debian community to run the survey and extend the model to understand from current contributors how the challenges antecede SVC and how SVC influences the intentions to leave the project.

Applying the research design to study other minorities (in different domains): Lack of diversity affects different OSS and other domains in STEM. While in this study

we focus on women in OSS, researchers can leverage this study's structure to investigate how the literature is positioned regarding the participation of other minority populations in OSS or even other domains.

Allen, K.-A., "Making sense of belonging", InPsych **41**, 3 (2019).

Allen, K.-A., *The Psychology of Belonging* (Routledge, 2020).

Anant, S. S., "Need to belong", Canadas Mental Health **14**, 2, 21–27 (1966).

Anant, S. S., "Belongingness and mental health: Some research findings", Acta psychologica **26**, 391–396 (1967).

Anant, S. S., "A cross-cultural study of belongingness, anxiety and self sufficiency", Acta psychologica **31**, 385–393 (1969).

Anderman, L. H., "Classroom goal orientation, school belonging and social goals as predictors of students' positive and negative affect following the transition to middle school.", Journal of Research & Development in Education (1999).

Anderman, L. H. and E. M. Anderman, "Social predictors of changes in students' achievement goal orientations", Contemporary educational psychology **24**, 1, 21–37 (1999).

Ardichvili, A., "Learning and knowledge sharing in virtual communities of practice: Motivators, barriers, and enablers", Advances in developing human resources **10**, 4, 541–554 (2008).

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

Atewologun, D., "Intersectionality theory and practice", in "Human Resource Management, Organizational Behavior, Research Methods, Social Issues", (Oxford Research Encyclopedia of Business and Management, 2018).

Auh, S., B. Menguc, S. Spyropoulou and F. Wang, "Service employee burnout and engagement: the moderating role of power distance orientation", J. Acad. Mark. Sci. **44**, 6 (2016).

Bacchelli, A. and C. Bird, "Expectations, outcomes, and challenges of modern code review", in "2013 35th International Conference on Software Engineering (ICSE)", pp. 712–721 (IEEE, 2013).

Balali, S., U. Annamalai, H. S. Padala, B. Trinkenreich, M. A. Gerosa, I. Steinmacher and A. Sarma, "Recommending tasks to newcomers in oss projects: How do mentors handle it?", in "Proceedings of the 16th International Symposium on Open Collaboration", pp. 1–14 (2020).

Balali, S., I. Steinmacher, U. Annamalai, A. Sarma and M. A. Gerosa, "Newcomers' barriers... is that all? an analysis of mentors' and newcomers' barriers in oss projects", Computer Supported Cooperative Work (CSCW) **27**, 3-6, 679–714 (2018).

Banaji, M. R. and N. Dasgupta, "The consciousness of social beliefs: A program of research on stereotyping and prejudice", Metacognition: Cognitive and social dimensions pp. 157–170 (1998).

Barcomb, A., K.-J. Stol, B. Fitzgerald and D. Riehle, "Managing episodic volunteers in free/libre/open source software communities", IEEE Transactions on Software Engineering (2020).

Barker, L. J., J. M. Cohoon and L. D. Thompson, "Work in progress—a practical model for achieving gender parity in undergraduate computing: Change the system, not the student", in "2010 IEEE Frontiers in Education Conference (FIE)", pp. S1H–1 (IEEE, 2010).

Barker, L. J., C. McDowell and K. Kalahar, "Exploring factors that influence computer science introductory course students to persist in the major", ACM Sigcse Bulletin **41**, 1, 153–157 (2009).

Bartlett, M., "Tests of significance in factor analysis.", Br J Psychol (1950).

Battistich, V., D. Solomon, D.-i. Kim, M. Watson and E. Schaps, "Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis", American educational research journal **32**, 3, 627–658 (1995).

Bauer, G. R., "Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity", Social science & medicine **110**, 10–17 (2014).

Baumeister, R. F. and M. R. Leary, "The need to belong: Desire for interpersonal attachments as a fundamental human motivation", Interpersonal development pp. 57–89 (2017).

Baumgartner, H. and J.-B. E. Steenkamp, "Response styles in marketing research: A cross-national investigation", Journal of marketing research **38**, 2, 143–156 (2001).

Beach, J., "Open source, open heart", XRDS: Crossroads, The ACM Magazine for Students **20**, 4, 24–27 (2014).

Becker, J.-M., A. Rai, C. Ringle and F. Völckner, "Discovering unobserved heterogeneity in structural equation models to avert validity threats", MISQ (2013).

Beecham, S., N. Baddoo, T. Hall, H. Robinson and H. Sharp, "Motivation in software engineering: A systematic literature review", Information and software technology **50**, 9-10, 860–878 (2008).

Bernard, H. R., *Research methods in anthropology: Qualitative and quantitative approaches* (Rowman & Littlefield, 2017).

Bernard, H. R. and C. C. Gravlee, *Handbook of methods in cultural anthropology* (Rowman & Littlefield, 2014).

Beyer, S., K. Rynes and S. Haller, "Deterrents to women taking computer science courses", IEEE technology and society magazine **23**, 1, 21–28 (2004).

Bitergia, "Gender-diversity analysis of the linux kernel technical contributions. accessed: 2020-10-16", https://blog.bitergia.com/2016/10/11/gender-diversity-analysis-of-the-linux-kernel-technical-contributions (2016).

Blanchard, A., D. A. Askay and K. A. Frear, "Sense of community in professional virtual communities", in "Virtual communities: Concepts, methodologies, tools and applications", pp. 1805–1820 (IGI Global, 2011).

Blanchard, A. L., "Developing a sense of virtual community measure", CyberPsychology & Behavior **10**, 6, 827–830 (2007).

Blanchard, A. L., "Testing a model of sense of virtual community", Computers in Human Behavior **24**, 5, 2107–2123 (2008).

Blanchard, A. L. and M. L. Markus, "Sense of virtual community-maintaining the experience of belonging", in "Proceedings of the 35th Annual Hawaii International Conference on System Sciences", pp. 3566–3575 (IEEE, 2002).

Blanchard, A. L. and M. L. Markus, "The experienced" sense" of a virtual community: Characteristics and processes", ACM Sigmis Database: the database for advances in information systems **35**, 1, 64–79 (2004).

Blincoe, K., O. Springer and M. R. Wrobel, "Perceptions of gender diversity's impact on mood in software development teams", IEEE Software **36**, 5, 51–56 (2019).

Bosu, A. and J. C. Carver, "Impact of peer code review on peer impression formation: A survey", in "2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement", pp. 133–142 (IEEE, 2013).

Bosu, A., J. C. Carver, C. Bird, J. Orbeck and C. Chockley, "Process aspects and social dynamics of contemporary code review: Insights from open source development and industrial practice at microsoft", IEEE Transactions on Software Engineering **43**, 1, 56–75 (2016).

Bosu, A. and K. Z. Sultana, "Diversity and inclusion in open source software (oss) projects: Where do we stand?", in "2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)", pp. 1–11 (IEEE, 2019).

Brown, R. and S. Pehrson, *Group processes: Dynamics within and between groups* (John Wiley & Sons, 2019).

Burnett, M., S. D. Fleming, S. Iqbal, G. Venolia, V. Rajaram, U. Farooq, V. Grigoreanu and M. Czerwinski, "Gender differences and programming environments: across programming populations", in "ESEM'10", pp. 1–10 (2010).

Burnett, M., S. Stumpf, J. Macbeth, S. Makri, L. Beckwith, I. Kwan, A. Peters and W. Jernigan, "Gendermag: A method for evaluating software's gender inclusiveness", Interacting with Computers **28**, 6, 760–787 (2016).

Burnett, M. M., L. Beckwith, S. Wiedenbeck, S. D. Fleming, J. Cao, T. H. Park, V. Grigoreanu and K. Rector, "Gender pluralism in problem-solving software", Interacting with computers **23**, 5, 450–460 (2011).

Burroughs, S. M. and L. T. Eby, "Psychological sense of community at work: A measurement system and explanatory framework", Journal of community psychology **26**, 6, 509–532 (1998).

Butler, J., "Gender is burning: Questions of appropriation and subversion", Cultural Politics **11**, 381–395 (1997).

Calvo, D., "The (in) visible barriers to free software: Inequalities in online communities in spain", Studies in Communication Sciences **21**, 1, 163–178 (2021).

Canedo, E. D., R. Bonifácio, M. V. Okimoto, A. Serebrenik, G. Pinto and E. Monteiro, "Work practices and perceptions from women core developers in oss communities", arXiv preprint arXiv:2007.13891 (2020).

Canfora, G., M. Di Penta, R. Oliveto and S. Panichella, "Who is going to mentor newcomers in open source projects?", in "Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering", p. 44 (ACM, 2012).

Carillo, K., S. Huff and B. Chawner, "What makes a good contributor? understanding contributor behavior within large free/open source software projects–a socialization perspective", The Journal of Strategic Information Systems **26**, 4, 322–359 (2017).

Catolino, G., F. Palomba, D. A. Tamburri, A. Serebrenik and F. Ferrucci, "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)", pp. 11–20 (IEEE, 2019).

Cazan, A.-M., E. Cocoradă and C. I. Maican, "Computer anxiety and attitudes towards the computer and the internet with romanian high-school and university students", Computers in Human Behavior **55**, 258–267 (2016).

Chang, H. H., P.-H. Hsieh and C. S. Fu, "The mediating role of sense of virtual community", Online Information Review (2016).

Chen, C.-W. and C.-S. Lin, "Building a sense of virtual community: The role of the features of social networking sites", Cyberpsychology, Behavior, and Social Networking **17**, 7, 460–465 (2014).

Chen, G.-L., S.-C. Yang and S.-M. Tang, "Sense of virtual community and knowledge contribution in a p3 virtual community: Motivation and experience", Internet Research (2013).

Cheryan, S., V. C. Plaut, P. G. Davies and C. M. Steele, "Ambient belonging: how stereotypical cues impact gender participation in computer science.", Journal of personality and social psychology **97**, 6, 1045 (2009).

Chih, W.-H., L.-C. Hsu and D.-K. Liou, "Understanding virtual community members' relationships from individual, group, and social influence perspectives", Industrial Management & Data Systems (2017).

Cho, Y., J. Park, S. Jeoung, B. Ju, J. You, A. Ju, C. K. Park, H. Y. Park *et al.*, "How do south korean female executives' definitions of career success differ from those of male executives?", European Journal of Training and Development (2017).

Cohen, J., *Statistical power analysis for the behavioral sciences* (Routledge, 2013).

Cohen, S., "Contextualizing toxicity in open source: a qualitative study", in "European Software Engineering Conference and Symposium on the Foundations of Software Engineering", (2021).

Cohoon, J. M., "Recruiting and retaining women in undergraduate computing majors", ACM SIGCSE Bulletin **34**, 2, 48–52 (2002).

Cole, E. R., "Intersectionality and research in psychology.", American psychologist **64**, 3, 170 (2009).

Corbet, J. and G. Kroah-Hartman, "Linux kernel development report", https://www.linuxfoundation.org/2017-linux-kernel-report-landing-page (2017).

Corbet, J., G. Kroah-Hartman and A. McPherson, "Linux kernel development. how fast it is going, who is doing it, what they are doing, and who is sponsoring it", The Linux Foundation (2012).

Cortina, K. S., S. Arel and J. P. Smith-Darden, "School belonging in different cultures: The effects of individualism and power distance", in "Frontiers in Education", vol. 2, p. 56 (Frontiers, 2017).

Creamer, E. G. and A. Laughlin, "Self-authorship and women's career decision making", Journal of college student development **46**, 1, 13–27 (2005).

da Silva, F. Q. and A. C. C. França, "Towards understanding the underlying structure of motivational factors for software engineers to guide the definition of motivational programs", Journal of Systems and Software **85**, 2, 216 – 226 (2012).

Dabbish, L., C. Stuart, J. Tsay and J. Herbsleb, "Social coding in github: transparency and collaboration in an open software repository", in "Proceedings of the ACM 2012 conference on computer supported cooperative work", pp. 1277–1286 (2012).

Dasgupta, N., "Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model", Psychological Inquiry **22**, 4, 231–246 (2011).

Dasgupta, N. and J. G. Stout, "Girls and women in science, technology, engineering, and mathematics: Steming the tide and broadening participation in stem careers", Policy Insights from the Behavioral and Brain Sciences **1**, 1, 21–29 (2014).

David, P. A. and J. S. Shapiro, "Community-based production of open-source software: What do we know about the developers who participate?", Information Economics and Policy **20**, 4, 364–398 (2008).

Davidson, J. L., R. Naik, U. A. Mannan, A. Azarbakht and C. Jensen, "On older adults in free/open source software: reflections of contributors and community leaders", in "2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)", pp. 93–100 (IEEE, 2014).

Dávila, M. C. and G. J. García, "Organizational identification and commitment: correlates of sense of belonging and affective commitment", The Spanish journal of psychology **15**, 1, 244–255 (2012).

de Gilder, D. and H. A. Wilke, "Expectation states theory and the motivational determinants of social influence", European review of social psychology **5**, 1, 243–269 (1994).

Deci, E. L. and R. M. Ryan, "Cognitive evaluation theory", in "Intrinsic motivation and self-determination in human behavior", (Springer, 1985).

Deci, E. L. and R. M. Ryan, "The support of autonomy and the control of behavior.", Journal of personality and social psychology **53**, 6, 1024 (1987).

Dias Canedo, E., H. Acco Tives, M. Bogo Marioti, F. Fagundes and J. A. Siqueira de Cerqueira, "Barriers faced by women in software development projects", Information **10**, 10, 309 (2019).

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G. G. Wagner, "Individual risk attitudes: Measurement, determinants, and behavioral consequences", Journal of the european economic association **9**, 3, 522–550 (2011).

Dominic, J., J. Houser, I. Steinmacher, C. Ritter and P. Rodeghero, "Conversational bot for newcomers onboarding to open source projects", in "Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops", pp. 46–50 (2020).

Dries, N., R. Pepermans and O. Carlier, "Career success: Constructing a multidimensional model", Journal of Vocational Behavior **73**, 2, 254–267 (2008).

Ducheneaut, N., "Socialization in an open source software community: A socio-technical analysis", Computer Supported Cooperative Work (CSCW) **14**, 4, 323–368 (2005).

Dwyer, S. C. and J. L. Buckle, "The space between: On being an insider-outsider in qualitative research", International journal of qualitative methods **8**, 1, 54–63 (2009).

Dyke, L. S. and S. A. Murphy, "How we define success: A qualitative study of what matters most to women and men", Sex Roles **55**, 5-6, 357–371 (2006).

Earley, C. P. and E. Mosakowski, "Creating hybrid team cultures: An empirical test of transnational team functioning", Academy of Management journal **43**, 1, 26–49 (2000).

Easterbrook, S., J. Singer, M.-A. Storey and D. Damian, "Selecting empirical methods for software engineering research", in "Guide to advanced empirical software engineering", pp. 285–311 (Springer, 2008).

Eghbal, N., "Roads and bridges", The Unseen labor behind our digital infrastructure (2016).

El Asri, I. and N. Kerzazi, "Where are females in oss projects? socio technical interactions", in "Working Conference on Virtual Enterprises", pp. 308–319 (Springer, 2019).

Ellonen, H.-K., M. Kosonen, A. Tarkiainen and L. Tonteri, "The positive outcomes of a sense of virtual community", International Journal of Web Based Communities **9**, 4, 465–482 (2013).

Espinosa, L., "Pipelines and pathways: Women of color in undergraduate stem majors and the college experiences that contribute to persistence", Harvard Educational Review **81**, 2, 209–241 (2011).

Evans, C. R., D. R. Williams, J.-P. Onnela and S. Subramanian, "A multilevel approach to modeling health inequalities at the intersection of multiple social identities", Social Science & Medicine **203**, 64–73 (2018).

Fagerholm, F., A. S. Guinea, J. Münch and J. Borenstein, "The role of mentoring and project characteristics for onboarding in open source software projects", in "Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement", p. 55 (ACM, 2014).

Fang, Y. and D. Neufeld, "Understanding sustained participation in open source software projects", Journal of Management Information Systems **25**, 4, 9–50 (2009).

Fatourou, P., Y. Papageorgiou and V. Petousi, "Women are needed in stem: European policies and incentives", Communications of the ACM **62**, 4, 52–52 (2019).

Faul, F., E. Erdfelder, A. Buchner and A.-G. Lang, "Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses", Behav Res Methods **41**, 4, 1149–1160 (2009).

Fehrenbacher, A. E. and D. Patel, "Translating the theory of intersectionality into quantitative and mixed methods for empirical gender transformative research on health", Culture, health & sexuality **22**, sup1, 145–160 (2020).

Felizardo, K. R., A. M. Ramos, C. d. O. Melo, É. F. de Souza, N. L. Vijaykumar and E. Y. Nakagawa, "Global and latin american female participation in evidence-based software engineering: a systematic mapping study", Journal of the Brazilian Computer Society **27**, 1, 1–22 (2021).

Feller, J. and B. Fitzgerald, "A framework analysis of the open source software development paradigm", in "ICIS 2000 proceedings of the twenty first international conference on information systems", pp. 58–69 (Association for Information Systems (AIS), 2000).

Fitzgerald, B., "The transformation of open source software", MIS quarterly **30**, 3, 587–598 (2006).

Fock, H., M. K. Hui, K. Au and M. H. Bond, "Moderation effects of power distance on the relationship between types of empowerment and employee satisfaction", J Cross Cult Psychol **44**, 2 (2013).

Ford, D., M. Behroozi, A. Serebrenik and C. Parnin, "Beyond the code itself: how programmers really look at pull requests", in "2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)", pp. 51–60 (IEEE, 2019).

Ford, D., A. Harkins and C. Parnin, "Someone like me: How does peer parity influence participation of women on stack overflow?", in "2017 IEEE symposium on visual languages and human-centric computing (VL/HCC)", pp. 239–243 (IEEE, 2017).

Forte, A. and C. Lampe, "Defining, understanding, and supporting open collaboration: Lessons from the literature", American Behavioral Scientist **57**, 5, 535–547 (2013).

Fossatti, M., "Gender, diversity, and inclusion in open source communities", XRDS: Crossroads, The ACM Magazine for Students **26**, 4, 46–48 (2020).

França, A. C. C., T. B. Gouveia, P. C. Santos, C. A. Santana and F. Q. da Silva, "Motivation in software engineering: A systematic review update", in "EASE 2011", pp. 154–163 (IET, 2011).

Frank, L. K., "Time perspectives", J. Soc. Phil. **4**, 293 (1938).

Freeman, T. M., L. H. Anderman and J. M. Jensen, "Sense of belonging in college freshmen at the classroom and campus levels", The Journal of Experimental Education **75**, 3, 203–220 (2007).

Frey, B. S., "On the relationship between intrinsic and extrinsic work motivation", International journal of industrial organization **4**, 15, 427–439 (1997).

Furtado, L., B. Cartaxo, C. Treude and G. Pinto, "How successful are open source contributions from countries with different levels of human development?", IEEE Software (2020).

Gan, R., L. S. Neo, J. Chin and M. Khader, "Change is the only constant: The evolving role of women in the islamic state in iraq and syria (isis)", Women & Criminal Justice **29**, 4-5, 204–220 (2019).

Garcia-Holgado, A., A. C. Diaz and F. J. Garcia-Penalvo, "Engaging women into stem in latin america: W-stem project", in "Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality", pp. 232–239 (2019).

Garrison, D., M. Cleveland-Innes, M. Koole and J. Kappelman, "Revisiting methodological issues in transcript analysis: Negotiated coding and reliability", The Internet and Higher Education **9**, 1, 1–8 (2006).

Gerlach, J., C.-G. Wu, L. Cunningham and C. Young, "An exploratory study of conflict over paying debian developers", Int. J. Open Source Softw. Process. **7**, 3 (2016).

Gerosa, M., I. Wiese, B. Trinkenreich, G. Link, G. Robles, C. Treude, I. Steinmacher and A. Sarma, "The shifting sands of motivation: Revisiting what drives contributors in open source", in "2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)", pp. 1046–1058 (IEEE, 2021).

Ghosh, R., R. Glott, B. Krieger and G. Robles, "Free/libre and open source software: Survey and study", (2002).

Gibbs, J. L., H. Kim and S. Ki, "Investigating the role of control and support mechanisms in members' sense of virtual community", Communication Research **46**, 1, 117–145 (2019).

Gila, A. R., J. Jaafa, M. Omar and M. Z. Tunio, "Impact of personality and gender diversity on software development teams' performance", in "2014 International Conference on Computer, Communications, and Control Technology (I4CT)", pp. 261–265 (IEEE, 2014).

Glaser, B. G. and A. L. Strauss, *Discovery of grounded theory: Strategies for qualitative research* (Routledge, 2017).

Glass, C. and A. Cook, "Leading at the top: Understanding women's challenges above the glass ceiling", The Leadership Quarterly **27**, 1, 51–63 (2016).

Good, C., A. Rattan and C. S. Dweck, "Why do women opt out? sense of belonging and women's representation in mathematics", Journal of personality and social psychology **102**, 4, 700 (2012).

Goodenow, C., "The psychological sense of school membership among adolescents: Scale development and educational correlates", Psychology in the Schools **30**, 1, 79–90 (1993).

Goodenow, C. and K. E. Grady, "The relationship of school belonging and friends' values to academic motivation among urban adolescent students", The Journal of Experimental Education **62**, 1, 60–71 (1993).

Greene, M. J., "On the inside looking in: Methodological insights and challenges in conducting qualitative insider research", The qualitative report **19**, 29, 1–13 (2014).

Greenwald, A. G. and M. R. Banaji, "Implicit social cognition: attitudes, self-esteem, and stereotypes.", Psychological review **102**, 1, 4 (1995).

Guizani, M., A. Chatterjee, B. Trinkenreich, M. E. May, G. J. Noa-Guevara, L. J. Russell, G. G. Cuevas Zambrano, D. Izquierdo-Cortazar, I. Steinmacher, M. A. Gerosa *et al.*, "The long road ahead: Ongoing challenges in contributing to large oss organizations and what to do", Proceedings of the ACM on Human-Computer Interaction **5**, CSCW2, 1–30 (2021).

Guizani, M., B. Trinkenreich, A. A. Castro-Guzman, I. Steinmacher, M. Gerosa and A. Sarma, "Perceptions of the state of d&i and d&i initiative in the asf", in "44th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)", (ACM, 2022).

Gusfield, J. R., *Community: A critical response* (Harper & Row New York, 1975).

Haas, L. and P. C. Hwang, "Is fatherhood becoming more visible at work? trends in corporate support for fathers taking parental leave in sweden.", Fathering: A Journal of Theory, Research & Practice about Men as Fathers **7**, 3 (2009).

Hagerty, B. M., J. Lynch-Sauer, K. L. Patusky, M. Bouwsema and P. Collier, "Sense of belonging: A vital mental health concept", Archives of psychiatric nursing **6**, 3, 172–177 (1992).

Hagerty, B. M. and K. Patusky, "Developing a measure of sense of belonging", Nursing research **44**, 1, 9–13 (1995).

Hair, J. F., J. J. Risher, M. Sarstedt and C. M. Ringle, "When to use and how to report the results of PLS-SEM", Eur. Bus. Rev. (2019).

Hair Jr, J. F., R. E. Anderson, R. L. Tatham and W. C. Black, "Multivariate data analysis with readings", (1995).

Hannah, D. R. and B. A. Lautsch, "Counting in qualitative research: Why to conduct it, when to avoid it, and when to closet it", Journal of Management Inquiry **20**, 1, 14–22 (2011).

Hannebauer, C. and V. Gruhn, "On the relationship between newcomer motivations and contribution barriers in open source projects", in "Proceedings of the 13th International Symposium on Open Collaboration", pp. 1–10 (2017).

Happe, L. and B. Buhnova, "Frustrations steering women away from software engineering", IEEE Software (2021).

Hars, A. and S. Ou, "Working for free motivations of participating in open source software projects", in "HICSS'04", pp. 25–31 (2004).

Hartzel, K., "How self-efficacy and gender issues affect software adoption and use", Communications of the ACM **46**, 9, 167–171 (2003).

Hattori, L. P. and M. Lanza, "On the nature of commits", in "2008 23rd IEEE/ACM International Conference on Automated Software Engineering-Workshops", pp. 63–71 (IEEE, 2008).

Haynes, C., N. M. Joseph, L. D. Patton, S. Stewart and E. L. Allen, "Toward an understanding of intersectionality methodology: A 30-year literature synthesis of black women's experiences in higher education", Review of Educational Research **90**, 6, 751–787 (2020).

Heilman, M. E., "Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder", Journal of social issues **57**, 4, 657–674 (2001).

Heilman, M. E., "Gender stereotypes and workplace bias", Research in organizational Behavior **32**, 113–135 (2012).

Henseler, J., C. M. Ringle and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling", J. Acad. Mark. Sci. **43**, 1, 115–135 (2015).

Henseler, J., C. M. Ringle and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing", in "New challenges to international marketing", (Emerald Group Publishing Limited, 2009).

Henshaw, A. L., "Where women rebel: Patterns of women's participation in armed rebel groups 1990–2008", International Feminist Journal of Politics **18**, 1, 39–60 (2016).

Hertel, G., S. Niedner and S. Herrmann, "Motivation of software developers in open source projects: an internet-based survey of contributors to the linux kernel", Research policy **32**, 7, 1159–1177 (2003).

Hofstede, G., *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (Sage publications, 2001).

Hofstede, G., "Dimensionalizing cultures: The hofstede model in context", Online readings in psychology and culture **2**, 1 (2011).

Homscheid, D., "The linux kernel project", in "Firm-Sponsored Developers in Open Source Software Projects", pp. 83–91 (Springer, 2020).

Homscheid, D., J. Kunegis and M. Schaarschmidt, "Private-collective innovation and open source software: Longitudinal insights from linux kernel development", in "Conference on e-Business, e-Services and e-Society", pp. 299–313 (Springer, 2015).

Hoyle, R. H. and A. M. Crawford, "Use of individual-level data to investigate group phenomena issues and strategies", Small Group Research **25**, 4, 464–485 (1994).

Huang, Y., K. Leach, Z. Sharafi, N. McKay, T. Santander and W. Weimer, "Biases and differences in code review using medical imaging and eye-tracking: genders, humans, and machines", in "Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering", pp. 456–468 (2020).

Huffman, A. H., J. Whetten and W. H. Huffman, "Using technology in higher education: The influence of gender roles on technology self-efficacy", Computers in Human Behavior **29**, 4, 1779–1786 (2013).

Hughes, M. M. and J. K. Dubrow, "Intersectionality and women's political empowerment worldwide", in "Measuring women's political empowerment across the globe", pp. 77–96 (Springer, 2018).

Hurtado, S. and D. F. Carter, "Effects of college transition and perceptions of the campus racial climate on latino college students' sense of belonging", Sociology of education pp. 324–345 (1997).

Imtiaz, N., J. Middleton, J. Chakraborty, N. Robson, G. Bai and E. Murphy-Hill, "Investigating the effects of gender bias on github", in "2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)", pp. 700–711 (IEEE, 2019).

Izquierdo, D., N. Huesman, A. Serebrenik and G. Robles, "Openstack gender diversity report", IEEE Software **36**, 1, 28–33 (2018).

Janoff-Bulman, R. and N. C. Carnes, "The model of moral motives", THE GUILFORD PRESS p. 223 (2018).

Jensen, C., S. King and V. Kuechler, "Joining free/open source software communities: An analysis of newbies' first interactions on project mailing lists", in "Proceedings of the 2011 44th Hawaii International Conference on System Sciences", HICSS '11, pp. 1–10 (IEEE Computer Society, Washington, DC, USA, 2011), URL http://dx.doi.org/10.1109/HICSS.2011.264.

Johnson, D. R., "Campus racial climate perceptions and overall sense of belonging among racially diverse women in stem majors", Journal of College Student Development **53**, 2, 336–346 (2012).

Johnson, D. R., M. Soldner, J. B. Leonard, P. Alvarez, K. K. Inkelas, H. T. Rowan-Kenyon and S. D. Longerbeam, "Examining sense of belonging among first-year undergraduates from different racial/ethnic groups", Journal of College Student Development **48**, 5, 525–542 (2007).

Kay, M., C. Matuszek and S. A. Munson, "Unequal representation and gender stereotypes in image search results for occupations", in "Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems", pp. 3819–3828 (2015).

Kazmi, A., "Women managers in different types of organisations: A representative research review", Journal of Entrepreneurship and Management **3**, 1 (2014).

Kerstetter, K., "Insider, outsider, or somewhere between: The impact of researchers' identities on the community-based research process", Journal of rural social sciences **27**, 2, 7 (2012).

Kim, E. and M. Drumwright, "Engaging consumers and building relationships in social media: How social relatedness influences intrinsic vs. extrinsic consumer motivation", Computers in Human Behavior **63**, 970–979 (2016).

Kim, J., H.-M. Kim and M. Kim, "The impact of a sense of virtual community on online community: does online privacy concern matter?", Internet Research (2020).

Kock, N., "Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM", International Journal of e-Collaboration **10**, 1 (2014).

Kofink, A., "Contributions of the under-appreciated: Gender bias in an open-source ecology", in "Companion Proceedings of the 2015 ACM SIGPLAN International Conference on Systems, Programming, Languages and Applications: Software for Humanity", pp. 83–84 (2015).

Koh, J., Y.-G. Kim and Y.-G. Kim, "Sense of virtual community: A conceptual framework and empirical validation", International journal of electronic commerce **8**, 2, 75–94 (2003).

Krause, J., W. Krause and P. Bränfors, "Women's participation in peace negotiations and the durability of peace", International Interactions **44**, 6, 985–1016 (2018).

Kuechler, V., C. Gilbertson and C. Jensen, "Gender differences in early free and open source software joining process", in "IFIP International Conference on Open Source Systems", pp. 78–93 (Springer, 2012).

Lakhani, K. R. and R. G. Wolf, "Why hackers do what they do: Understanding motivation and effort in free/open source software projects", (2003).

Lee, A. and J. C. Carver, "Floss participants' perceptions about gender and inclusiveness: a survey", in "2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)", pp. 677–687 (IEEE, 2019).

Lee, J. F., "Gender representation in japanese efl textbooks–a corpus study", Gender and Education **30**, 3, 379–395 (2018).

Lee, K.-Y., "Factors affecting members' sense of belonging in virtual community", Journal of the Korean Operations Research and Management Science Society **35**, 1, 19–45 (2010).

Lewin, K., "A dynamic theory of personality: Selected papers", The Journal of Nervous and Mental Disease **84**, 5, 612–613 (1936).

Li, Z., W. Seering, T. Tao and S. Cao, "Understanding community behaviors in for-profit open source hardware projects", in "Proceedings of the Design Society: International Conference on Engineering Design", pp. 2397–2406 (Cambridge University Press, 2019).

Lim, S., "Job satisfaction of information technology workers in academic libraries", Library & Information Science Research **30**, 2, 115–121 (2008).

Lin, B., G. Robles and A. Serebrenik, "Developer turnover in global, industrial open source projects: Insights from applying survival analysis", in "2017 IEEE 12th International Conference on Global Software Engineering (ICGSE)", pp. 66–75 (IEEE, 2017).

Lin, B. and A. Serebrenik, "Recognizing gender of stack overflow users", in "Proceedings of the 13th International Conference on Mining Software Repositories", pp. 425–429 (2016).

Lin, H.-F., "Determinants of successful virtual communities: Contributions from system characteristics and social factors", Information & Management **45**, 8, 522–527 (2008).

Louis, K. S., A. C. Barton *et al.*, "Tales from the science education crypt: A critical reflection of positionality, subjectivity, and reflexivity in research", in "Forum Qualitative Sozialforschung/Forum: Qualitative Social Research", vol. 3 (2002).

MacKinnon, C. A., "Intersectionality as method: A note", Signs: Journal of Women in Culture and Society **38**, 4, 1019–1030 (2013).

Maher, F. A. and M. K. T. Tetreault, *The feminist classroom: Dynamics of gender, race, and privilege* (Rowman & Littlefield Publishers, 2001).

Maheshwari, G., "A review of literature on women's leadership in higher education in developed countries and in vietnam: Barriers and enablers", Educational Management Administration & Leadership p. 17411432211021418 (2021).

Makarova, E. and W. Herzog, "Trapped in the gender stereotype? the image of science among secondary school students and teachers", Equality, Diversity and Inclusion: An International Journal (2015).

Mani, A. and R. Mukherjee, "A study of foss 2013 survey data using clustering techniques", in "2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)", pp. 118–121 (IEEE, 2016).

Marcoulides, G. A. and C. Saunders, "Editor's comments: PLS: a silver bullet?", MISQ pp. iii–ix (2006).

Margolis, J., R. Estrella, J. Goode, J. J. Holme and K. Nao, *Stuck in the shallow end: Education, race, and computing* (MIT press, 2017).

Margolis, J. and A. Fisher, *Unlocking the clubhouse: Women in computing* (MIT press, 2002).

Martínez-Torres, M. R. and M. C. Díaz-Fernández, "Current issues and research trends on open-source software communities", Technology Analysis & Strategic Management **26**, 1, 55–68 (2014).

Maslow, A. H., "A theory of human motivation.", Psychological review **50**, 4, 370 (1943).

McClintock, C. G., "Social motivation—a set of propositions", Behav. Sci. **17**, 5 (1972).

McCullough, L., "Women's leadership in science, technology, engineering and mathematics: Barriers to participation", in "Forum on Public Policy Online", vol. 2011 (ERIC, 2011).

McDougall, W., *An introduction to social psychology* (Psychology Press, 2015).

McGowan, D., "Legal implications of open-source software", U. Ill. L. Rev. (2001).

McMillan, D. W. and D. M. Chavis, "Sense of community: A definition and theory", Journal of community psychology **14**, 1, 6–23 (1986).

Melaku, T. M., A. Beeman, D. G. Smith and W. B. Johnson, "Be a better ally", Harvard Business Review **98**, 6, 135–139 (2020).

Mendez, C., H. S. Padala, Z. Steine-Hanson, C. Hilderbrand, A. Horvath, C. Hill, L. Simpson, N. Patil, A. Sarma and M. Burnett, "Open source barriers to entry, revisited: A sociotechnical perspective", in "Proceedings of the 40th International Conference on Software Engineering", pp. 1004–1015 (2018a).

Mendez, C., A. Sarma and M. Burnett, "Gender in open source software: what the tools tell", in "Proceedings of the 1st International Workshop on Gender Equality in Software Engineering", pp. 21–24 (2018b).

Menezes, Á. and R. Prikladnicki, "Diversity in software engineering", in "Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering", pp. 45–48 (2018).

Merriam, S. B., J. Johnson-Bailey, M.-Y. Lee, Y. Kee, G. Ntseane and M. Muhamad, "Power and positionality: Negotiating insider/outsider status within and across cultures", International journal of lifelong education **20**, 5, 405–416 (2001).

Merriam, S. B. and E. J. Tisdell, *Qualitative research: A guide to design and implementation* (John Wiley & Sons, 2015).

Miles, M. B. and A. M. Huberman, *Qualitative data analysis: An expanded sourcebook* (sage, 1994).

Miller, C., S. Cohen, D. Klug, B. Vasilescu and C. Kästner, "Did you miss my comment or what? understanding toxicity in open source discussions", in "Int'l Conf. Software Engineering", (2022).

Molleman, E. and J. Slomp, "The impact of team and work characteristics on team functioning", Human Factors and Ergonomics in Manufacturing & Service Industries **16**, 1, 1–15 (2006).

Moon, E., "Gendered patterns of politeness in free/libre open source software development", in "2013 46th Hawaii International Conference on System Sciences", pp. 3168–3177 (IEEE, 2013).

Moreno, L., Y. González and P. Martínez, "Women in computer science: Survey on the perception of the women's participation in stem studies", in "Proceedings of the XV International Conference on Human Computer Interaction", pp. 1–2 (2014).

Morrison, P., R. Pandita, E. Murphy-Hill and A. McLaughlin, "Veteran developers' contributions and motivations: An open source perspective", in "2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)", pp. 171–179 (IEEE, 2016).

Mullangi, S. and R. Jagsi, "Imposter syndrome: treat the cause, not the symptom", Jama **322**, 5, 403–404 (2019).

Murakami, Y., M. Tsunoda and H. Uwano, "Wap: Does reviewer age affect code review performance?", in "2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE)", pp. 164–169 (IEEE, 2017).

Nafus, D., "'patches don't have gender': What is not open in open source software", New Media & Society **14**, 4, 669–683 (2012).

Nakakoji, K., Y. Yamamoto, Y. Nishinaka, K. Kishida and Y. Ye, "Evolution patterns of open-source software systems and communities", in "Proceedings of the international workshop on Principles of software evolution", pp. 76–85 (ACM, 2002).

Neel, R., D. T. Kenrick, A. E. White and S. L. Neuberg, "Individual differences in fundamental social motives", J Pers Soc Psychol **110**, 6 (2016).

Oreg, S. and O. Nov, "Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values", Computers in human behavior **24**, 5, 2055–2073 (2008).

Ortu, M., G. Destefanis, S. Counsell, S. Swift, R. Tonelli and M. Marchesi, "How diverse is your team? investigating gender and nationality diversity in github teams", Journal of Software Engineering Research and Development **5**, 1, 1–18 (2017).

Overflow, S., "Stack overflow annual developer survey", URL https://insights.stackoverflow.com/survey/2018 (2018).

Padala, S. H., C. J. Mendez, L. F. Dias, I. Steinmacher, Z. S. Hanson, C. Hilderbrand, A. Horvath, C. Hill, L. D. Simpson, M. Burnett *et al.*, "How gender-biased tools shape newcomer experiences in oss projects", IEEE Transactions on Software Engineering (2020).

Pantic, K. and J. Clarke-Midura, "Factors that influence retention of women in the computer science major: A systematic literature review", Journal of Women and Minorities in Science and Engineering **25**, 2 (2019).

Parker, S. W., "Opening up to open source", Richmond Journal of Law & Technology **6**, 5, 24 (2000).

Paul, R., A. Bosu and K. Z. Sultana, "Expressions of sentiments during code reviews: Male vs. female", in "2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER)", pp. 26–37 (IEEE, 2019).

Peters, L., "Managing software professionals", in "IEMC'03 Proceedings. Managing Technologically Driven Organizations: The Human Side of Innovation and Change", pp. 61–66 (IEEE, 2003).

Pinto, G., I. Steinmacher and M. A. Gerosa, "More common than you think: An in-depth study of casual contributors", in "2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)", vol. 1, pp. 112–123 (IEEE, 2016).

Pinto, G. H. L., F. Figueira Filho, I. Steinmacher and M. A. Gerosa, "Training software engineers using open-source software: the professors' perspective", in "2017 IEEE 30th Conference on Software Engineering Education and Training (CSEE&T)", pp. 117–121 (IEEE, 2017).

Pittman, L. D. and A. Richmond, "University belonging, friendship quality, and psychological adjustment during the transition to college", The Journal of Experimental Education **76**, 4, 343–362 (2008).

Podsakoff, P. M., S. B. MacKenzie, J.-Y. Lee and N. P. Podsakoff, "Common method biases in behavioral research: a critical review of the literature and recommended remedies.", Journal of applied psychology **88**, 5, 879 (2003).

Porter, A. M., "Physics phds ten years later: Success factors and barriers in career paths. results from the phd plus 10 study.", AIP Statistical Research Center (2019).

Powell, W. E., D. S. Hunsinger and B. D. Medlin, "Gender differences within the open source community: An exploratory study", Journal of Information Technology **21**, 4, 29–37 (2010).

Prana, G. A. A., D. Ford, A. Rastogi, D. Lo, R. Purandare and N. Nagappan, "Including everyone, everywhere: Understanding opportunities and challenges of geographic gender-inclusion in oss", IEEE Transactions on Software Engineering (2021).

Qiu, H. S., Y. L. Li, S. Padala, A. Sarma and B. Vasilescu, "The signals that potential contributors look for when choosing open-source projects", Proceedings of the ACM on Human-Computer Interaction **3**, CSCW, 1–29 (2019a).

Qiu, H. S., A. Nolte, A. Brown, A. Serebrenik and B. Vasilescu, "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)", pp. 688–699 (IEEE, 2019b).

Qiu, H. S., A. Nolte, A. Brown, A. Serebrenik and B. Vasilescu, "Going farther together: The impact of social capital on sustained participation in open source", in "Proceedings of the 41st International Conference on Software Engineering", ICSE '19, pp. 688–699 (IEEE Press, Piscataway, NJ, USA, 2019c), URL https://doi.org/10.1109/ICSE.2019.00078.

Qiu, Y., K. J. Stewart and K. M. Bartol, "Joining and socialization in open source women's groups: an exploratory study of kde-women", in "IFIP International Conference on Open Source Systems", pp. 239–251 (Springer, 2010).

Ramakrishnan, A., D. Sambuco and R. Jagsi, "Women's participation in the medical profession: insights from experiences in japan, scandinavia, russia, and eastern europe", Journal of Women's Health **23**, 11, 927–934 (2014).

Raman, N., M. Cao, Y. Tsvetkov, C. Kästner and B. Vasilescu, "Stress and burnout in open source: Toward finding, understanding, and mitigating unhealthy interactions", in "ICSE", (2020).

Ramos, D. and R. Wright-Mair, "Imposter syndrome: A buzzword with damaging consequences. diverse issues in higher education", https://www.diverseeducation.com/tenure/article/15109066/imposter-syndrome-a-buzzword-with-damaging-consequences (2021).

Rastogi, A., N. Nagappan and G. Gousios, "Geographical bias in github: Perceptions and reality", Tech. rep., DSpace (2016).

Rastogi, A., N. Nagappan, G. Gousios and A. van der Hoek, "Relationship between geographical location and evaluation of developer contributions in github", in "Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement", pp. 1–8 (2018).

Rheingold, H., *The virtual community, revised edition: Homesteading on the electronic frontier* (MIT press, 2000).

Riehle, D., P. Riemer, C. Kolassa and M. Schmidt, "Paid vs. volunteer work in open source", in "2014 47th Hawaii International Conference on System Sciences", pp. 3286–3295 (IEEE, 2014).

Ringle, C., D. Da Silva and D. Bido, "Structural equation modeling with the smartpls", Bido, D., da Silva, D., & Ringle, C.(2014). Structural Equation Modeling with the Smartpls. Brazilian Journal Of Marketing **13**, 2 (2015).

Roberts, J. A., I.-H. Hann and S. A. Slaughter, "Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects", Management science **52**, 7, 984–999 (2006).

Robles, G., L. Arjona Reina, A. Serebrenik, B. Vasilescu and J. M. González-Barahona, "Floss 2013: A survey dataset about free software contributors: challenges for curating, sharing, and combining", in "Proceedings of the 11th Working Conference on Mining Software Repositories", pp. 396–399 (2014).

Robles, G., L. A. Reina, J. M. González-Barahona and S. D. Domínguez, "Women in free/libre/open source software: The situation in the 2010s", in "IFIP International Conference on Open Source Systems", pp. 163–173 (Springer, 2016).

Robles, G., I. Steinmacher, P. Adams and C. Treude, "Twenty years of open source software: From skepticism to mainstream", IEEE Software **36**, 6, 12–15 (2019).

Robson, N., "Diversity and decorum in open source communities", in "Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering", pp. 986–987 (2018).

Rodríguez-Pérez, G., R. Nadri and M. Nagappan, "Perceived diversity in software engineering: a systematic literature review", Empirical Software Engineering **26**, 5, 1–38 (2021).

Russo, D. and K.-J. Stol, "Pls-sem for software engineering research: An introduction and survey", ACM Computing Surveys (CSUR) **54**, 4, 1–38 (2021).

Ryan, R. and E. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.", Am. Psychol. **55**, 1, 68 (2000).

Sach, R., H. Sharp and M. Petre, "Software engineers' perceptions of factors in motivation: The work, people, obstacles", in "2011 International Symposium on Empirical Software Engineering and Measurement", pp. 368–371 (IEEE, 2011).

Santos, F., B. Trinkenreich, J. F. Pimentel, I. Wiese, I. Steinmacher, A. Sarma and M. A. Gerosa, "How to choose a task? mismatches in perspectives of newcomers and existing contributors", in "Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement", pp. 114–124 (2022).

Santos, F., I. Wiese, B. Trinkenreich, I. Steinmacher, A. Sarma and M. Gerosa, "Can i solve it? identifying apis required to complete oss task", in "2021 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR)", (2021).

Sarstedt, M., J.-M. Becker, C. Ringle and M. Schwaiger, "Uncovering and treating unobserved heterogeneity with fimix-pls: which model selection criterion provides an appropriate number of segments?", Schmalenbach Business Review **63**, 1 (2011).

Sarstedt, M. and J.-H. Cheah, "Partial least squares structural equation modeling using smartPLS: a software review", (2019).

Sarstedt, M., C. M. Ringle and J. F. Hair, "Treating unobserved heterogeneity in PLS-SEM: A multi-method approach", in "Partial least squares path modeling", pp. 197–217 (Springer, 2017).

Sax, L. J., J. M. Blaney, K. J. Lehman, S. L. Rodriguez, K. L. George and C. Zavala, "Sense of belonging in computing: The role of introductory courses for women and underrepresented minority students", Social Sciences **7**, 8, 122 (2018).

Schaarschmidt, M. and K.-J. Stol, "Company soldiers and gone-natives: role conflict and career ambition among firm-employed open source developers", in "Proceedings of the International Conference on Information Systems", (Association for Information Systems, 2018).

Secules, S., C. McCall, J. A. Mejia, C. Beebe, A. S. Masters, M. L. Sánchez-Peña and M. Svyantek, "Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community", Journal of Engineering Education **110**, 1, 19–43 (2021).

Sharan, F., "Asf committer diversity survey. accessed: 2020-10-16", https://cwiki.apache.org/confluence/display/COMDEV/ASF+Committer+Diversity+Survey+-+2016 (2016).

Sharma, P. N., S. L. Daniel, T. R. Chung and V. Grover, "A motivation-hygiene model of open source software code contribution and growth", J Assoc Inform Sys **23**, 1, 165–195 (2022).

Silva, J., I. Wiese, D. M. German, C. Treude, M. A. Gerosa and I. Steinmacher, "A theory of the engagement in open source projects via summer of code programs", in "Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering", pp. 421–431 (2020a).

Silva, J. D. O., I. S. Wiese, D. M. German, I. F. Steinmacher and M. A. Gerosa, "How long and how much: What to expect from summer of code participants?", in "2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)", pp. 69–79 (IEEE, 2017).

Silva, J. O., I. Wiese, D. M. German, C. Treude, M. A. Gerosa and I. Steinmacher, "Google summer of code: Student motivations and contributions", Journal of Systems and Software **162**, 110487 (2020b).

Silveira, K. K. and R. Prikladnicki, "A systematic mapping study of diversity in software engineering: a perspective from the agile methodologies", in "2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)", pp. 7–10 (IEEE, 2019).

Singh, A., V. Bhadauria, A. Jain and A. Gurung, "Role of gender, self-efficacy, anxiety and testing formats in learning spreadsheets", Computers in Human Behavior **29**, 3, 739–746 (2013).

Singh, V., "Women-only spaces of open source", in "2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)", pp. 17–20 (IEEE Computer Society, 2019a).

Singh, V., "Women participation in open source software communities", in "Proceedings of the 13th European Conference on Software Architecture-Volume 2", pp. 94–99 (2019b).

Singh, V. and W. Brandon, "Open source software community inclusion initiatives to support women participation", in "IFIP International Conference on Open Source Systems", pp. 68–79 (Springer, 2019).

Solomon, D., M. Watson, V. Battistich, E. Schaps and K. Delucchi, "Creating classrooms that students experience as communities", American Journal of Community Psychology **24**, 6, 719–748 (1996).

Spencer, D., *Card sorting: Designing usable categories* (Rosenfeld Media, 2009).

Spencer, S. J., C. M. Steele and D. M. Quinn, "Stereotype threat and women's math performance", Journal of experimental social psychology **35**, 1, 4–28 (1999).

Spichkova, M., H. Schmidt and C. Trubiani, "Role of women in software architecture: an attempt at a systematic literature review", in "Proceedings of the 11th European Conference on Software Architecture: Companion Proceedings", pp. 31–34 (2017).

Spinellis, D., "Open source and professional advancement", IEEE Software **23**, 5, 70–71 (2006).

Steinmacher, I., S. Balali, B. Trinkenreich, M. Guizani, D. Izquierdo-Cortazar, G. G. Cuevas Zambrano, M. A. Gerosa and A. Sarma, "Being a mentor in open source projects", Journal of Internet Services and Applications **12**, 1, 1–33 (2021).

Steinmacher, I., A. P. Chaves, T. Conte and M. A. Gerosa, "Preliminary empirical identification of barriers faced by newcomers to open source software projects.", in "Proceedings of the 28th Brazilian Symposium on Software Engineering", SBES '14, pp. 51–60 (IEEE Computer Society, 2014a).

Steinmacher, I., T. Conte and M. A. Gerosa, "Understanding and supporting the choice of an appropriate task to start with in open source software communities", in "2015 48th Hawaii International Conference on System Sciences", HICSS '15, pp. 5299–5308 (IEEE, 2015a).

Steinmacher, I., T. Conte, M. A. Gerosa and D. Redmiles, "Social barriers faced by newcomers placing their first contribution in open source software projects", in "Proceedings of the 18th ACM conference on Computer supported cooperative work & social computing", pp. 1379–1392 (ACM, 2015b).

Steinmacher, I., T. U. Conte, C. Treude and M. A. Gerosa, "Overcoming open source project entry barriers with a portal for newcomers", in "Proceedings of the 38th International Conference on Software Engineering", pp. 273–284 (2016).

Steinmacher, I., M. A. Gerosa and D. Redmiles, "Attracting, onboarding, and retaining newcomer developers in open source software projects", in "Workshop on GSD in a CSCW Perspective", (2014b).

Steinmacher, I., G. Robles, B. Fitzgerald and A. Wasserman, "Free and open source software development: the end of the teenage years", Journal of Internet Services and Applications **8**, 1, 17, URL https://doi.org/10.1186/s13174-017-0069-9 (2017).

Steinmacher, I., M. A. G. Silva and M. A. Gerosa, "Barriers faced by newcomers to open source projects: a systematic review", in "IFIP International Conference on Open Source Systems", pp. 153–163 (Springer, 2014c).

Steinmacher, I., I. S. Wiese, T. Conte, M. A. Gerosa and D. Redmiles, "The hard life of open source software project newcomers", in "Proceedings of the 7th international workshop on cooperative and human aspects of software engineering", pp. 72–78 (ACM, 2014d).

Stewart, K. and S. Gosain, "The impact of ideology on effectiveness in open source software development teams", MISQ **30**, 2 (2006).

Stol, K.-J. and B. Fitzgerald, "Inner source–adopting open source development practices in organizations: a tutorial", IEEE Software **32**, 4, 60–67 (2014).

Stol, K.-J. and B. Fitzgerald, "The abc of software engineering research", ACM Transactions on Software Engineering and Methodology (TOSEM) **27**, 3, 11 (2018).

Stone, M., "Cross-validatory choice and assessment of statistical predictions", J R Stat Soc Series B Stat Methodol **36**, 2, 111–133 (1974).

Strauss, A. and J. M. Corbin, *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory* (SAGE Publications, 2007), 3rd edn.

Strayhorn, T. L., "Exploring the impact of facebook and myspace use on first-year students' sense of belonging and persistence decisions", Journal of College Student Development **53**, 6, 783–796 (2012).

Strayhorn, T. L., *College students' sense of belonging: A key to educational success for all students* (Routledge, 2018).

Sutanto, J., A. Kankanhalli and B. C. Y. Tan, "Eliciting a sense of virtual community among knowledge contributors", ACM Transactions on Management Information Systems (TMIS) **2**, 3, 1–17 (2011).

Tajfel, H. and J. C. Turner, "The social identity theory of intergroup behavior", in "Polit Psychol", pp. 276–293 (Psychology Press, 2004).

Tamburri, D. A., R. Kazman and H. Fahimi, "The architect's role in community shepherding", IEEE Software **33**, 6, 70–79 (2016).

Tamilmani, K., N. Rana, N. Prakasam and Y. Dwivedi, "The battle of brain vs. heart: A literature review and meta-analysis of "hedonic motivation" use in utaut2", Int J Inf Manage **46** (2019).

Tan, X., M. Zhou and B. Fitzgerald, "Scaling open source communities: An empirical study of the linux kernel", in "International Conference on Software Engineering", (IEEE, 2020).

Taylor, J. and R. Dantu, "For love or money? examining reasons behind oss developers' contributions", Inform Sys Manage **39**, 2, 122–137 (2022).

Tenenhaus, M., V. E. Vinzi, Y.-M. Chatelin and C. Lauro, "Pls path modeling", Computational statistics & data analysis **48**, 1 (2005).

Terrell, J., A. Kofink, J. Middleton, C. Rainear, E. Murphy-Hill, C. Parnin and J. Stallings, "Gender differences and bias in open source: Pull request acceptance of women versus men", PeerJ Computer Science **3**, e111 (2017).

Thoman, D. B., J. A. Arizaga, J. L. Smith, T. S. Story and G. Soncuya, "The grass is greener in non-science, technology, engineering, and math classes: Examining the role of competing belonging to undergraduate women's vulnerability to being pulled away from science", Psychology of Women Quarterly **38**, 2, 246–258 (2014).

Tomasetto, C., F. R. Alparone and M. Cadinu, "Girls' math performance under stereotype threat: the moderating role of mothers' gender stereotypes.", Developmental psychology **47**, 4, 943 (2011).

Tonteri, L., M. Kosonen, H.-K. Ellonen and A. Tarkiainen, "Antecedents of an experienced sense of virtual community", Computers in Human Behavior **27**, 6, 2215–2223 (2011).

Tourani, P., B. Adams and A. Serebrenik, "Code of conduct in open source projects", in "2017 IEEE 24th international conference on software analysis, evolution and reengineering (SANER)", pp. 24–33 (IEEE, 2017).

Trainer, E. H., C. Chaihirunkarn, A. Kalyanasundaram and J. D. Herbsleb, "Community code engagements: summer of code & hackathons for community building in scientific software", in "Proceedings of the 18th International Conference on Supporting Group Work", pp. 111–121 (2014).

Trainer, E. H., C. Chaihirunkarn, A. Kalyanasundaram and J. D. Herbsleb, "From personal tool to community resource: What's the extra work and who will do it?", in "18th ACM Conference on Computer Supported Cooperative Work & Social Computing", pp. 417–430 (2015).

Trinkenreich, B., "Replication package", https://figshare.com/s/f8df9d2449a116920305 (2022).

Trinkenreich, B., R. Britto, M. Gerosa and I. Steinmacher, "An empirical investigation on the challenges faced by women in the software industry: A case study", in "44th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)", (ACM, 2022a).

Trinkenreich, B., M. A. Gerosa and I. Steinmacher, "Women in open source: We need to talk about it", Computer **55**, 12, 145–149 (2022b).

Trinkenreich, B., M. Guizani, I. Wiese, T. Conte, M. Gerosa, A. Sarma and I. Steinmacher, "The pot of gold at the end of the rainbow: What is success for open source contributors?", arXiv preprint arXiv:2105.08789 (2021a).

Trinkenreich, B., M. Guizani, I. Wiese, A. Sarma and I. Steinmacher, "Hidden figures: Roles and pathways of successful oss contributors", Computer Supported Cooperative Work (CSCW) **4**, 180 (2020a).

Trinkenreich, B., M. Guizani, I. Wiese, A. Sarma and I. Steinmacher, "Hidden figures: Roles and pathways of successful oss contributors", Computer Supported Cooperative Work (CSCW) **4**, 180 (2020b).

Trinkenreich, B., M. Guizani, I. S. Wiese, T. Conte, M. Gerosa, A. Sarma and I. Steinmacher, "Pots of gold at the end of the rainbow: What is success for open source contributors", IEEE Transactions on Software Engineering pp. 1–1 (2021b).

Trinkenreich, B., K.-J. Stol, A. Sarma, D. German, M. Gerosa and I. Steinmacher, "Do i belong? modeling sense of virtual community among linux kernel contributors", in "2023 IEEE/ACM 45rd International Conference on Software Engineering (ICSE)", (IEEE, 2023).

Trinkenreich, B., I. Wiese, A. Sarma, M. Gerosa and I. Steinmacher, "Women's participation in open source software: A survey of the literature", arXiv preprint arXiv:2105.08777 (2021c).

Turkle, S., *The Second Self: Computers and the Human Spirit* (Simon & Schuster, Inc., New York, NY, USA, 2005a).

Turkle, S., *The second self: Computers and the human spirit* (Mit Press, 2005b).

Van Someren, M. W., Y. F. Barnard and J. A. Sandberg, "The think aloud method: a practical approach to modelling cognitive", London: AcademicPress **11** (1994).

Vasilescu, B., A. Capiluppi and A. Serebrenik, "Gender, representation and online participation: A quantitative study", Interacting with Computers **26**, 5, 488–511 (2014).

Vasilescu, B., V. Filkov and A. Serebrenik, "Perceptions of diversity on git hub: A user survey", in "2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering", pp. 50–56 (IEEE, 2015a).

Vasilescu, B., D. Posnett, B. Ray, M. G. van den Brand, A. Serebrenik, P. Devanbu and V. Filkov, "Gender and tenure diversity in github teams", in "Proceedings of the 33rd annual ACM conference on human factors in computing systems", pp. 3789–3798 (2015b).

Vasilescu, B., A. Serebrenik and V. Filkov, "A data set for social diversity studies of github teams", in "2015 IEEE/ACM 12th working conference on mining software repositories", pp. 514–517 (IEEE, 2015c).

Vedres, B. and O. Vasarhelyi, "Gendered behavior as a disadvantage in open source software development", EPJ Data Science **8**, 1, 25 (2019).

Veilleux, N., R. Bates, D. Jones, C. Allendoerfer and J. Crawford, "The role of belonging in engagement, retention and persistence in computer science", in "Proceedings of the 43rd ACM technical symposium on Computer Science Education", pp. 707–707 (2012).

Vendome, C., M. Linares-Vasquez, G. Bavota, M. Di Penta, D. M. German and D. Poshyvanyk, "When and why developers adopt and change software licenses", in "2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)", pp. 31–40 (IEEE, 2015).

Visagie, S. and E. Koekemoer, "What it means to succeed: Personal perceptions of career success held by senior managers", South African Journal of Business Management **45**, 1, 43–54 (2014).

Von Krogh, G., S. Haefliger, S. Spaeth and M. W. Wallin, "Carrots and rainbows: Motivation and social practice in open source software development", MIS quarterly pp. 649–676 (2012).

Von Krogh, G., S. Spaeth and K. R. Lakhani, "Community, joining, and specialization in open source software innovation: a case study", Research policy **32**, 7, 1217–1241 (2003).

Wang, Z., Y. Wang and D. Redmiles, "Competence-confidence gap: A threat to female developers' contribution on github", in "2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)", pp. 81–90 (IEEE, 2018).

Weijters, B. and H. Baumgartner, "Misresponse to reversed and negated items in surveys: A review", Journal of Marketing Research **49**, 5, 737–747 (2012).

Wohlin, C. and A. Aurum, "Towards a decision-making structure for selecting a research design in empirical software engineering", Empirical Software Engineering **20**, 6, 1427–1455 (2015).

Wu, D. J., J. Park and N. Dasgupta, "The influence of male faces on stereotype activation among women in stem: An erp investigation", Biological Psychology **156**, 107948 (2020).

Wu, Y., Y. Manabe, T. Kanda, D. M. German and K. Inoue, "A method to detect license inconsistencies in large-scale open source projects", in "2015 IEEE/ACM 12th Working Conference on Mining Software Repositories", pp. 324–333 (IEEE, 2015).

Wu, Y., Y. Manabe, T. Kanda, D. M. German and K. Inoue, "Analysis of license inconsistency in large collections of open source projects", Empirical Software Engineering **22**, 3, 1194–1222 (2017).

Wurzelová, P., F. Palomba and A. Bacchelli, "Characterizing women (not) contributing to open-source", in "2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)", pp. 5–8 (IEEE, 2019).

Xu, B. and D. Li, "An empirical study of the motivations for content contribution and community participation in wikipedia", Information & management **52**, 3, 275–286 (2015).

Xu, J., S. Christley and G. Madey, "Application of social network analysis to the study of open source software", in "The economics of open source software development", pp. 247–269 (Elsevier, 2006).

Ye, Y. and K. Kishida, "Toward an understanding of the motivation of open source software developers", in "25th International Conference on Software Engineering, 2003. Proceedings.", pp. 419–429 (IEEE, 2003).

Zacchiroli, S., "Gender differences in public code contributions: a 50-year perspective", IEEE Software (2020).

Zhao, L., Y. Lu, B. Wang, P. Y. Chau and L. Zhang, "Cultivating the sense of belonging and motivating user participation in virtual communities: A social capital perspective", International Journal of Information Management **32**, 6, 574–588 (2012).

Zhou, M. and A. Mockus, "Who will stay in the floss community? modeling participant's initial behavior", IEEE Transactions on Software Engineering **41**, 1, 82–99 (2014).

Zlotnick, F., "Github open source survey 2017", URL https://doi.org/10.5281/zenodo.806811 (2017a).

Zlotnick, F., "Github open source survey 2017", http://opensourcesurvey.org/2017/ (2017b).