The Shifting Sands of Toxicity: The Evolving Nature of Interpersonal Challenges in Open Source

Sarthak Bharadwaj, Fabio Santos, Bianca Trinkenreich
Colorado State University, Fort Collins, USA
{sarthak.bharadwaj, fabio.deabreusantos, bianca.trinkenreich}@colostate.edu

***Abstract*—Background: The sustainability of Open Source Software (OSS) projects relies on attracting and retaining contributors. Interpersonal challenges, whether experienced or witnessed, can discourage participation, alter behavior, or drive contributors away. Aims: This study examines how interpersonal challenges in OSS communities persist over time and explores their behavioral consequences on contributors' decisions and actions. Method: We analyze data from two large GitHub Open Source surveys conducted in 2017 (n=5,495) and 2024 (n=8,452), evaluating changes in reported interpersonal challenges (RQ1) and differential consequences of exposure between 2017 and 2024 (RQ2). Results: Our findings reveal a significant increase in reported interpersonal challenges in 2024 compared to 2017. Contributors more frequently reported severe challenges such as threats of violence, impersonation, sustained harassment, stalking, and doxxing. The behavioral impact has shifted: experiencing rudeness, stalking, and name-calling became strongly linked to stopping contributions, adopting pseudonyms, working privately, and avoiding offline events. Witnessing harmful behaviors like name-calling and impersonation also became stronger predictors of working privately or advocating for Codes of Conduct. These trends show toxicity is not only more pervasive but increasingly damaging to OSS participation and community health. Conclusions: Results highlight a concerning rise in interpersonal challenges within OSS communities, with rudeness emerging as the most impactful. The growing influence of toxic behaviors on contributors' decisions to withdraw, conceal identities, isolate collaboration, and avoid offline engagement underscores the urgent need for stronger, proactive community support. Sustaining healthy OSS projects requires both technical excellence and deliberate investment in social infrastructure to foster respectful collaboration spaces.**

***Index Terms*—OSS, Open Source, GitHub, Interpersonal Challenges, Toxicity**

I. Introduction

Open Source Software (OSS) communities thrive on global collaboration, shared learning, and collective innovation. However, these ecosystems are not immune to the challenges of negative interpersonal interactions [1, 2]. As OSS projects grow in size and visibility, so does the volume and complexity of communication among contributors. In the context of productive dialogue, interpersonal challenges represent a very critical barrier to contributing to OSS [3]. Toxic behaviors, such as harassment, name-calling, doxxing, and other insults, can emerge that threaten the contributor's well-being and inclusion, and may impact the project's sustainability [1, 4]. The consequences of such toxicity are far-reaching: developers report stress, burnout, exclusion, and, in some cases, a complete withdrawal from participation.

979-8-3315-9147-2/25/$31.00 ©2025 IEEE

The literature has extensively examined the diverse challenges faced by contributors in Open Source Software (OSS) communities, ranging from social and onboarding difficulties to project specific dynamics and interpersonal relationships [2, 5–8]. Over the years, several initiatives have been introduced to reduce toxicity in OSS projects, one prominent example being the adoption of Codes of Conduct by many projects [9, 10]. Despite these efforts, there remains a significant gap in understanding how toxicity in OSS has evolved over time, whether it has decreased, persisted, or even worsened.

This brings us to our first research question:

RQ1: How have interpersonal challenges in OSS shifted from 2017 to 2024?

People respond to online toxicity in diverse ways [4, 11]. Prior research highlights that when toxic behaviors arise in OSS communities, it is crucial for community leaders to mitigate the harm and for members to be informed and prepared to respond appropriately [12]. However, existing systems and practices often fall short of resolving these issues, placing a disproportionate burden on maintainers [4]. Recent studies have begun to quantitatively explore the broader consequences of these experiences in OSS communities. Such challenges have been associated with contributors feeling unwelcome or alienated within these environments [13]. Uncivil comments during code reviews, for example, have been shown to lead to a variety of outcomes, including conflict escalation, discontinued discussions, shifts toward technical clarification, or invocation of a code of conduct [14]. The consequences of interpersonal challenges can include various behaviors, such as contributors reducing or stopping their contributions, concealing their presence or identity, or speaking out to raise awareness within the community. However, what remains underexplored is a nuanced understanding of how these interpersonal challenges impact contributor participation and how this impact has changed over time. This motivates our second research question:

RQ2: How have different interpersonal challenges impacted OSS contributors' behaviors, and how have these impacts shifted from 2017 to 2024?

The current absence of a comprehensive understanding that delineates the impact of challenges on a broader set of contributor behaviors and participation implies that endeavors by OSS communities to retain contributors are likely to fail.

Our study closes this gap by analyzing interpersonal challenges in OSS using two large-scale GitHub OSS surveys conducted in 2017 [15] and 2024 [16]. Our replication package is available at [17].

Our study contributes to the software engineering literature as a human-centered, empirically grounded investigation into the evolving socio-technical dynamics affecting OSS contributors. It follows a sample survey strategy [18] and represents a descriptive, knowledge-seeking study that examines interpersonal challenges in OSS communities over time. The primary beneficiaries are human stakeholders, particularly OSS contributors and community managers, with the main contribution being empirical insights into behavioral responses to toxicity [19].

Specifically, we started by analyzing trends in reported interpersonal challenges between 2017 and 2024 (RQ1). Afterward, we investigated how different forms of interpersonal challenges impact contributors' participation in OSS projects and how this impact shifted between 2017 and 2024 (RQ2).

Our results reveal a significant uprise from 2017 to 2024 in reports of both experiencing and witnessing the majority of interpersonal challenges in OSS communities.

Our data also indicate that experiencing some of the interpersonal challenges increasingly influenced contributors' decisions to stop participating in OSS projects, adopt pseudonyms, work privately, alter their online presence, engage in private and public discussions about these issues, and make offline changes such as avoiding conferences. Witnessing such challenges had a growing impact on decisions to stop contributing, use pseudonyms, work privately, and advocate for the adoption of a code of conduct.

By examining how specific interpersonal challenges have evolved over time and how they impact contributors' behaviors and OSS participation, our study offers a nuanced, time-based understanding of the toxic interactions between contributors in OSS communities. This longitudinal perspective not only advances the academic discourse in behavioral software engineering, but also provides practical insights for sustaining healthy OSS ecosystems. Our findings highlight which aspects of the contributor experience are most closely tied to participation and retention, offering actionable guidance for community leaders seeking to foster long-term engagement, innovation, and project sustainability. Moreover, the insights gained from our work may extend beyond OSS, informing efforts across the broader software industry to cultivate more inclusive, equitable, and supportive work environments.

II. Related Work

Interpersonal challenges in OSS. Open source has long carried a reputation for being an aggressive and unwelcoming environment [20]. Toxicity and incivility are widespread in OSS discussions and code reviews, negatively impacting contributor engagement, collaboration, and the overall health of communities [4, 12]. These issues are particularly pronounced for contributors from underrepresented groups, who often face even more hostile and exclusionary collaboration environments. Harassing language, sexual jokes, insults, and bullying remain disturbingly common [21, 22]. Singh and Brandon [23] report that fewer than 5% of online communities can be considered "safe" for women contributors, free from sexism and discrimination. Despite these challenges, a quantitative analysis of 355 OSS package websites found that only 10% include a code of conduct or similar governance policies [24].

Consequences of Interpersonal challenges in OSS. Recent research [13] explored how various interpersonal challenges, such as doxxing, sexual harassment, stereotyping, and unwelcoming language, affect OSS contributors' sense of welcomeness. Using a large-scale survey conducted by the Linux Foundation, the study confirmed a strong negative association between experiencing interpersonal challenges and feeling welcome and highlighted the serious consequences of interpersonal challenges on participation and retention in OSS. Our study adds to this body of knowledge by investigating more consequences of experiencing (and also witnessing) interpersonal challenges in OSS.

III. Research Design

Answering RQ1 involved using descriptive statistics and Bonferroni-adjusted Chi-Square tests, while RQ2 involved using Likelihood Ratio Tests followed by Logistic Regressions corrected with Bonferroni test.

We answer our research questions using data from two GitHub Open Source Surveys conducted in 2017 [15] and 2024 [16]. These surveys were designed and administered by GitHub, and data are publicly available [15, 16].

Eligibility for participating in the surveys was determined based on user activity, specifically visits to three distinct projects or three interactions within a single project within a 30-minute time frame. Invitations were shown for up to three subsequent page views or until dismissed by the user. The introductory text on the survey's landing page informed respondents that anonymous results would be publicly released as an open dataset, emphasized that all questions were optional, and provided instructions for accessing translated versions of the survey (available in Traditional Chinese, Japanese, Spanish, and Russian). We used questions that were repeated in the 2017 and 2024 questionnaires.

Interpersonal Challenges: Respondents were presented with the following list of 10 interpersonal challenges and asked two questions: whether they had personally experienced each challenge and whether they had witnessed it happening to others.

(1) *Rudeness* often involves a failure to acknowledge social norms of politeness and civility, such as interrupting, using offensive language, or disregarding others' feelings.[12]

(2) *Name Calling* is also known as identity attacks and refers to the use of derogatory or insulting language directed at an individual or group, often with the intention of belittling or demeaning. [25]

(3) *Threats of violence* are statements or actions that imply harm or negative consequences will occur to an individual or group if they do not comply with certain demands. [25]

(4) *Impersonation* occurs when an individual or group falsely assumes or uses the identity of another person.[26]

(5) *Sustained Harassment* refers to repeated and ongoing acts of intimidation, bullying, or harmful behavior directed at an individual, often over an extended period of time.

(6) *Cross-Platform Harassment* involves the use of multiple online platforms or communication channels to harass or target an individual, and may include sending harmful messages through social media, emails, online forums, and messaging apps.

(7) *Stalking* refers to actions meant to intimidate or instill fear [27, 28].

(8) *Unsolicited sexual advances or comments* refer to unsolicited or inappropriate gestures or actions with sexual intent, directed at an individual, often causing discomfort, harassment, or distress.[13]

(9) *Stereotyping* represents discrimination based on perceived demographic characteristics, usually referring to a form of negative and fixed impression that can happen implicitly or explicitly and relate to a socially shared set of beliefs about the traits of members of a social category.[29]

(10) *Doxxing*, derived from "*document dropping*," involves the public disclosure of private or personal information about an individual or organization, often with harmful intentions [30].

Behavioral Consequences of Interpersonal Challenges: Both surveys included a question asking what, if any, further actions contributors took as a result of experiencing or witnessing such challenges. The following behavioral consequences were presented: (1) stopping contributions to a project, (2) contributing under a pseudonym, (3) increasing the use of private channels for work, questions, or collaboration, (4) changing or deleting a username, (5) altering or removing content from their public online presence, (6) suggesting the creation or revision of a code of conduct, engaging in (7) public or (8) private discussions with community members about the issue, or (9) making offline changes such as avoiding meetups or conferences.

A. RQ1 Analysis

To address RQ1, we conducted Chi-square and Bonferroni tests to assess whether there were statistically significant changes between 2017 and 2024 in the proportions of contributors who reported experiencing or witnessing each type of interpersonal challenge. For each challenge, we compared responses that indicate whether there is experience or observation over the years.

We complemented this analysis with descriptive statistics, examining percentages of contributors reporting each challenge in both years. This allowed us to verify the direction of change, whether reports increased or decreased, for cases where significant differences were found.

B. RQ2 Analysis

To address RQ2, we created 36 logistic regression models, having the dependent variables each of the consequences (stopped contributing to a project; started contributing under a pseudonym; worked privately or collaborated in private channels more often; changed or deleted a username; removed or changed content on my public online presence; suggested the creation or modification of a code of conduct; engaged in public discussion with community members about the issue; engaged in private discussion with community members about the issue; and made changes in offline life, e.g. stopped attending meetups or conferences, etc.)

Independent variables were binary responses to the survey on whether the contributor experienced or witnessed each of the interpersonal challenges (rudeness; name calling; threats of violence; impersonation; harassment over a sustained period; harassment across multiple platforms; stalking; unsolicited sexual advances or comments; stereotyping based on perceived demographic characteristics; malicious publication of personal information-doxxing).

We considered only complete responses, that is, survey entries containing valid values for all variables analyzed, and excluded incomplete responses without applying any imputation methods for missing data. To assess the quality of our logistic regression models and ensure the validity of our findings, we have checked our variables' covariance matrix and VIF to confirm the absence of multicollinearity.

As our model tested 36 different consequences, we applied the Bonferroni correction to adjust the significance level and control for multiple comparisons. This helped reduce the risk of Type I errors, which could have arisen due to the large number of statistical tests conducted, ensuring that the conclusions drawn from these tests were robust and reliable.

To examine whether the effects of negative experiences on various consequences have changed between 2017 and 2024, we conducted logistic regression analyses using a combined dataset that includes responses from both years. We created a binary indicator variable `year`, coded as 0 for 2017 and 1 for 2024, to distinguish between the two time periods. For each binary consequence variable, such as `stopped contributing`, `use pseudonyms`, `suggest COC`, and other reported outcomes, we first fit a logistic regression model that included all one-hot encoded negative experience variables, the `year` indicator, and interaction terms between each negative experience and the year. This full model allows us to test whether the relationship between negative experiences and consequences differs across years. Next, we fit a reduced model that includes only the main effects of the negative experiences and the `year` variable, excluding interaction terms. To compare the two models, we applied a likelihood ratio test (LRT), which evaluates whether the inclusion of the interaction terms significantly improves model fit. A statistically significant (p-value < 0.05) from the LRT indicates that the full model fits the data significantly better than the reduced model, suggesting that the influence of negative experiences on the specific consequence has changed over time. This modeling approach enables us to capture and assess evolving patterns in how negative experiences impact user behavior and participation across the two survey years.

The results of these likelihood ratio tests for consequences due to witnessing negative events are presented in Table III. We found that the interaction terms significantly improved the model fit for several consequences, indicating meaningful differences in the effect of witnessing negative experiences between 2017 and 2024. Specifically, the full model was significantly better than the reduced model for the outcomes *Stopped Contribution*, *Used Pseudonyms*, *Worked in Private*, and *Suggested Code of Conduct*. These findings suggest that the influence of witnessing negative behavior on these specific community-related actions has shifted over time. In contrast, the consequences such as *Changed Username*, *Changed Online Presence*, *Private Community Discussions*, *Public Community Discussions*, and *Offline Changes* did not benefit significantly from the inclusion of interaction terms. This implies that the behavioral patterns in response to witnessing negative events have remained relatively stable between 2017 and 2024 for these specific outcomes.

The results of the likelihood ratio tests for consequences due to directly experiencing negative events are presented in Table II. Our findings reveal that several behavioral outcomes have significantly changed over time in response to experienced negativity. The full logistic regression model, which includes interaction terms with the year variable, fit the data significantly better than the reduced model for a number of consequences. These include *Stopped Contribution*, *Used Pseudonyms*, *Worked in Private*, *Changed Online Presence*, *Private Community Discussions*, *Public Community Discussions*, and *Offline Changes*, suggesting that the behavioral responses to negative experiences have evolved substantially from 2017 to 2024. In contrast, consequences such as *Changed Username* and *Suggested Code of Conduct* did not show significant improvements in model fit with the interaction terms, indicating that these particular responses have remained relatively consistent across the two time periods examined.

To interpret the models, we compared the change in odds ratio of the same regressions from 2017 and 2024.

IV. Analyses and Results

In this section, we evaluate each research question according to Section III.

A. How have interpersonal challenges in OSS shifted from 2017 to 2024?

In this section, we present our findings for RQ1, which examine differences in how frequently respondents experienced and witnessed interpersonal challenges in 2017 and 2024. Using Chi-square and Bonferroni tests, we assessed whether the reporting of *experienced* and *witnessed* challenges changed significantly over the years.

Our results in Table I indicate that reports of most types of interpersonal challenges experienced changed significantly between 2017 and 2024. However, *rudeness* and *name-calling* did *not* show statistically significant differences in reporting rates between the two years. For witnessed challenges, most types also showed significant shifts over time, *except for rudeness, name-calling, and sexual advances*, which did *not* exhibit meaningful changes.

We used descriptive analysis to evaluate differences in the percentages of reported challenges over time.

As shown in Figure 1, the percentages of reported *experienced* interpersonal challenges increased across the board in 2024, suggesting that contributors are facing all types of interpersonal challenges MORE FREQUENTLY over time.

TABLE I: Chi-square results comparing interpersonal challenges experienced (vs. not experienced) and witnessed (vs. not witnessed) in 2017 vs. 2024. Significant changes are indicated by * and **, * indicates $p < 0.05$, and ** indicates $p < 0.01$

Experienced		
Interpersonal Challenges	**Chi-square Value**	**Bonferroni p-value**
Threats of Violence	20.1362	0.00**
Impersonation	16.9544	0.00**
Sustained Harassment	19.3111	0.00**
Cross-platform Harassment	23.8454	0.00**
Stalking	21.5671	0.00**
Doxxing	27.3999	0.00**
Rudeness	06.9564	0.09
Name-calling	04.7497	0.32
Stereotyping	10.0126	0.02*
Sexual Advances	08.2605	0.04*
Witnessed		
Interpersonal Challenges	**Chi-square Value**	**Bonferroni p-value**
Threats of Violence	11.5510	0.00**
Impersonation	29.8223	0.00**
Sustained Harassment	25.9468	0.00**
Cross-platform Harassment	46.1210	0.00**
Stalking	30.8159	0.00**
Doxxing	45.7178	0.00**
Rudeness	00.5646	1.00
Name-calling	00.0000	1.00
Stereotyping	08.0533	0.05*
Sexual Advances	00.6191	1.00

When comparing reports of *witnessed* interpersonal challenges, we observe an overall increase in 2024 across nearly all challenge types. This trend is visible even for those challenges that did not show statistically significant differences, indicating a broader perception of interpersonal difficulties within the community.

To gain a deeper understanding of how each form of interpersonal challenge affected contributors in 2017 and 2024, we performed a descriptive statistical analysis. We plotted a comparative bar graph showing the percentage of contributors who reported witnessing or experiencing each form of negative behavior in both years. This visualization highlights the changes in prevalence over time, offering a clear comparison of how the reporting of witnessed or experienced interpersonal challenges has evolved.

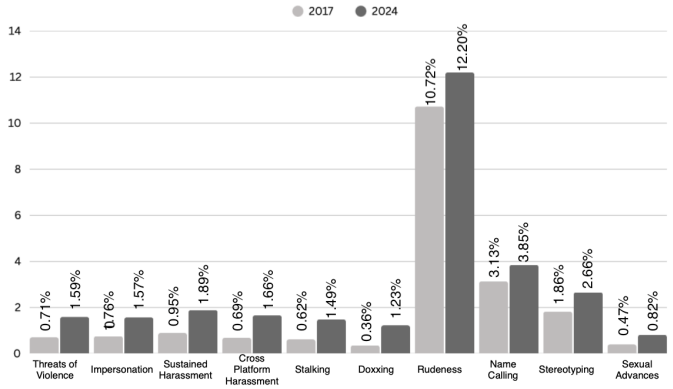

Fig. 1: Percentages of respondents who reported *experiencing* each type of interpersonal challenge on 2017 and 2024

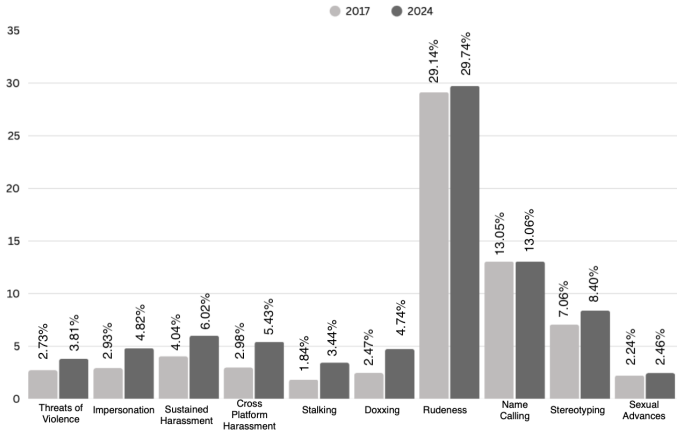

Fig. 2: Percentages of respondents who reported *witnessing* each type of interpersonal challenge on 2017 and 2024

While the total number of complete survey responses from the GitHub dataset increased from 5,495 in 2017 to 8,452 in 2024, the percentage of respondents who experienced and witnessed any kind of interpersonal challenge also increased over time. Specifically, the percentage of OSS contributors who reported witnessing interpersonal challenges rose from 32.17% (n=1,768) in 2017 to 34.42% (n=2,909) in 2024. Similarly, the percentage of contributors who reported personally experiencing such challenges increased from 12.30% (n=676) to 15.10% (n=1,277) over the same period.

We individually examined the evolution of interpersonal challenges experienced (see Figures 1 and 2) over the two years and noticed that the percentage of contributors experiencing and witnessing each form of interpersonal challenges increased from 2017 to 2024.

Our analysis reveals notable increases in forms of interpersonal challenges reported by contributors in the OSS community. Experienced challenges, threats of violence increased from 0.71% (n=39) in 2017 to 1.59% (n=134) in 2024, while impersonation rose from 0.76% (n=42) to 1.57% (n=133). Reports of sustained harassment nearly doubled from 0.95% (n=52) to 1.89% (n=160), and doxxing increased more than threefold from 0.36% (n=20) to 1.23% (n=104). These patterns point to an intensification of more severe forms of online toxicity in OSS spaces over time.

For Witnessed incidents reports of threats of violence rose from 2.73% (n=150) in 2017 to 3.81% (n=322) in 2024. Impersonation increased from 2.93% (n=161) to 4.82% (n=407), and sustained harassment from 4.04% (n=222) to 6.02% (n=509). Similarly, cases of cross-platform harassment rose from 2.98% (n=164) to 5.43% (n=459), and doxxing nearly doubled from 2.47% (n=136) to 4.74% (n=401).

RQ1. Our results revealed a significant increase in reports of experiencing interpersonal challenges in OSS communities between 2017 and 2024. Contributors reported experiencing threats of violence, impersonation, sustained and cross-platform harassment, stalking, doxxing, stereotyping, and sexual advances more frequently in 2024 than in 2017. Reports of witnessing these challenges, excluding sexual advances, showed a significant rise in most cases. In contrast, reports of experiencing or witnessing rudeness and name-calling did not show statistically significant changes over the years.

B. How have different interpersonal challenges impacted OSS contributors' behaviors, and how have these impacts shifted from 2017 to 2024?

In this section, we present our findings for RQ2, which examine differences in the impact of *experiencing* and *witnessing* interpersonal challenges in contributors' behaviors and participation, including their decisions to stop contributing, using a pseudonym, using private channels, changing username, changing online presence, suggesting a code of conduct, engaging in public or private discussions about the issue, or making offline changes such as avoiding meetups or conferences. For the sake of simplicity, and following the survey question that asked about those behaviors as *"results of experiencing or witnessing"* the challenges, we call them consequences.

1) The shifting impact of interpersonal challenges across the years: We conducted Likelihood Ratio tests to determine whether there had been a significant shift in the effects of experiencing or witnessing interpersonal challenges on various forms of consequences between 2017 and 2024.

In Table II, our results revealed a significant change over the years in the impact of experiencing interpersonal challenges on STOP CONTRIBUTING, USING A PSEUDONYM, WORKING IN PRIVATE, CHANGING ONLINE PRESENCE, DISCUSSING THE EXPERIENCED ISSUE BOTH IN PRIVATE AND IN PUBLIC, and

TABLE II: Likelihood Ratio Test Results for Consequence due to Experiencing Interpersonal Challenges. * denotes $p < 0.05$ is considered statistically significant.

Consequence	Likelihood Ratio Statistic (G^2)	p-value
Stopped Contribution	93.6811	0.00*
Used Pseudonyms	32.8372	0.00*
Worked In Private	23.0841	0.01*
Changed Username	09.0246	0.53
Changed Online Presence	18.8870	0.04*
Suggested Code of Conduct	10.9939	0.36
Private Community Discussions	49.0461	0.00*
Public Community Discussions	36.6628	0.00*
Offline Changes	18.8579	0.04*

making OFFLINE CHANGES (e.g., not going to conferences anymore).

No significant changes were detected in the impact of experiencing interpersonal challenges between 2017 and 2024 on the behaviors of CHANGING USERNAME, SUGGESTING A CODE OF CONDUCT.

TABLE III: Likelihood Ratio Test Results for Consequence due to Witnessing Interpersonal Challenges. * denotes $p < 0.05$ is considered statistically significant.

Consequence	Likelihood Ratio Statistic (G^2)	p-value
Stopped Contribution	19.6251	0.03*
Used Pseudonyms	21.7233	0.02*
Worked In Private	41.2904	0.00*
Changed Username	13.1673	0.21
Changed Online Presence	15.6627	0.11
Suggested Code of Conduct	21.0439	0.02*
Private Community Discussions	11.2650	0.34
Public Community Discussions	15.3290	0.12
Offline Changes	08.7291	0.56

Similarly, Table III shows a significant shift in the impact of witnessing interpersonal challenges over time. This shift affected three of the same behaviors influenced by direct experience (STOPPING CONTRIBUTIONS, USING A PSEUDONYM, and WORKING PRIVATELY), and additionally, SUGGESTING OF A CODE OF CONDUCT.

No significant changes were detected on the impact of interpersonal challenges between 2017 and 2024 on the behaviors of CHANGING USERNAME, CHANGING ONLINE PRESENCE, DISCUSSING THE EXPERIENCED ISSUE IN PRIVATE OR IN PUBLIC, and OFFLINE CHANGES.

Given these findings, where many consequences exhibited significant changes in the effects of witnessing or experiencing interpersonal challenges, we conducted regression analysis for each consequence to further investigate how the influence of experiencing or witnessing each interpersonal challenge had evolved between 2017 and 2024.

The following analysis provides a more detailed examination of the impact of interpersonal challenges on individual consequences in 2017 and 2024.

2) The impacts of experiencing interpersonal challenges over the years: To investigate the impact of experiencing different interpersonal challenges on each consequence over the years, as described in Section III, we modeled a regression model for each consequence in 2017 and 2024, having the experienced challenges as predictors (see Table IV). The columns shaded in light gray highlight the consequences for which the influence of experiencing interpersonal challenges changed significantly between the two years.

We now discuss the results of the consequences that showed a significant impact from the experienced interpersonal challenges over the years (as per Table II).

STOP CONTRIBUTING: In 2017, contributors who reported experiencing rudeness had 9.8 times higher odds of stopping contributions compared to those who did not. By 2024, this association intensified, with those experiencing rudeness being 46 times more likely to stop contributing. That year, experiencing name-calling and stereotyping also significantly increased the odds of stopping contributions, whereas in 2017, these challenges were not significantly associated with this behavior.

USING PSEUDONYM: In 2017, contributors who experienced impersonation, name-calling, and stereotyping had significantly higher odds of using a pseudonym than in 2024, compared with those who did not experience the same challenges. By 2024, the effect of impersonation remained significantly associated with those challenges, though reduced. In 2024, stalking and rudeness emerged as strong predictors of using pseudonyms, while stereotyping was no longer significantly associated with this behavior.

WORK PRIVATELY: In 2017, experiencing Rudeness was the only challenge significantly associated with working privately. By 2024, besides Rudeness, this behavior was significantly influenced by experiencing a variety of other interpersonal challenges: impersonation, stalking, stereotyping and name-calling, suggesting a broader range of interpersonal challenges were driving contributors to isolate their work. Rudeness stood out in 2024; contributors who experienced rudeness had 11.7 times higher odds of working privately compared with those who did not experience.

CHANGE ONLINE PRESENCE: In 2017, two experienced challenges significantly influenced changes in online presence, rudeness, and stereotyping. By 2024, stalking and name-calling were also predictors of this behavior, while rudeness stood out again. Contributors who experienced rudeness had 10.6 times higher odds of changing online presence compared with those who did not experience.

ENGAGE IN PRIVATE DISCUSSIONS: While it may seem intuitive that those who experienced interpersonal challenges would be more likely to engage in private discussions about them, our findings still highlight notable trends in how this behavior evolved over time. In 2017, contributors who experienced rudeness and stereotyping were significantly more likely to discuss these issues privately. By 2024, the likelihood of private discussions increased substantially, with contributors who experienced rudeness being 36 times more likely to engage in such discussions compared to those who did not experience rudeness.

ENGAGE IN PUBLIC DISCUSSIONS: In 2017, contributors who experienced rudeness were significantly more inclined to

TABLE IV: The Impact of Experiencing Interpersonal Challenges on Contributors' Behaviors (2017 and 2024)

	Stop Contributing	Pseudonym	Work Privately	Change Username	Change Online Presence	Suggest COC	Private Discussions	Public Discussions	Offline Changes
Threats of Violence	-11.15	-90.46	69.50	11.93	168.00	233.13	5.74	130.73	195.01
Impersonation	-24.74	1753.22**	0.98	127.11	7.17	229.47	44.47	-25.58	20.71
Sustained Harassment	3.71	-42.49	-36.25	2.53	-66.37	80.05	47.16	48.42	-25.99
Cross-Platform Harassment	19.34	-91.03	140.40	21.44	217.47	4.61	138.75	39.24	7.80
Stalking	50.34	382.41	49.32	174.37	170.16	-60.52	-34.55	-13.64	107.01
Doxxing	46.74	118.51	83.75	75.25	318.30	-44.60	115.58	27.61	105.18
Rudeness	986.17**	6.63	1035.31**	153.08	375.06**	462.95**	762.70**	888.13**	473.24**
Name-calling	10.57	445.91*	41.72	184.21	110.75	50.74	54.63	43.13	66.89
Stereotyping	66.99	822.50**	130.65	48.34	351.11**	153.95	292.08 **	107.36	160.72
Sexual Advances	32.41	505.03	46.63	330.36	-37.48	310.60	77.24	106.94	47.01
2017 R^2	*0.158*	*0.215*	*0.204*	*0.140*	*0.205*	*0.175*	*0.200*	*0.189*	*0.158*
Threats of Violence	-51.06	38.08	26.49	40.93	-48.36	17.37	-13.64	1.95	-41.12
Impersonation	13.40	456.41**	157.21**	346.15**	81.13	153.73	10.28	85.89	-69.44
Sustained Harassment	-25.59	-53.72	112.02	-36.99	60.13	34.05	116.29*	108.30	75.05
Cross-Platform Harassment	-13.49	-29.36	-41.09	-19.55	-11.37	181.45	57.78	27.81	-30.21
Stalking	152.08*	208.62*	306.50**	615.83**	330.43**	61.92	-0.97	21.13	191.06
Doxxing	-4.62	165.19	12.42	71.32	197.75*	-15.17	-23.41	64.12	443.23**
Rudeness	4,639.35**	671.04**	1,174.94**	494.04**	1,063.83**	698.95**	3,640.80**	3,786.88**	1,231.10**
Name-calling	80.66**	221.81**	193.67**	99.71	201.49**	65.93	96.13**	58.25	152.09**
Stereotyping	179.06**	93.65	137.88**	114.43	201.54**	64.87	117.37**	79.98	196.57**
Sexual Advances	-3.78	-57.36	306.15	-21.58	-53.13	267.10*	131.78	-7.74	46.83
2024 - R^2	*0.410*	*0.291*	*0.342*	*0.294*	*0.349*	*0.305*	*0.405*	*0.389*	*0.340*

* denotes $p < 0.05$ (Bonferroni-corrected, statistically significant), ** denotes $p <= 0.01$ (highly statistically significant).
The light gray columns indicate consequences for which the impact of experiencing interpersonal challenges significantly changed between 2017 and 2024.

discuss these issues in public forums. By 2024, those who experienced rudeness had 37 times higher odds of engaging in public discussions compared to those who did not experience it.

Taken together, these findings reveal a dual trend in how contributors respond to interpersonal challenges: both private and public discussions have become increasingly common over time. While it is expected that those who experience harm may seek out trusted peers for private conversations, the dramatic rise in odds, from already significant levels in 2017 to extreme values in 2024 (e.g., odds = 3,640.80 for private and odds = 3,786.88 for public discussions following rudeness), underscores a growing need for both confidential support and communal accountability.

This suggests that contributors are not only turning inward to process harmful experiences through informal, peer-based networks but are also increasingly willing to speak out publicly, challenging problematic behavior and advocating for change. The simultaneous rise in both private and public discourse highlights a shift in community norms toward greater recognition, validation, and visibility of interpersonal harm in OSS environments.

MAKE OFFLINE CHANGES: In 2017, contributors who experienced rudeness had significantly higher odds of making offline changes, such as avoiding in-person conferences, compared to those who did not. By 2024, this trend intensified: contributors who experienced rudeness were 12 times more likely to make such changes. Additional challenges also became significant predictors, including doxxing (4 times higher odds), name-calling (152% higher odds), and stereotyping (196% higher odds). These findings suggest a growing impact of online interpersonal harm on contributors' real-world participation and visibility within the community.

3) The impacts of witnessing interpersonal challenges over the years: To investigate the impact of witnessing different interpersonal challenges on each consequence over the years, as described in Section III, we modeled a regression model for each consequence in 2017 and 2024, now having the witnessed challenges as predictors (see Table V). The columns shaded in light gray highlight the consequences for which the influence of witnessing interpersonal challenges changed significantly between the two years.

We now discuss the results of the consequences that showed a significant impact from the witnessed interpersonal challenges over the years (as per Table III).

STOP CONTRIBUTING: In 2017, contributors who reported witnessing rudeness had 25.3 times higher odds of stopping contribution compared to those who did not. By 2024, the effect of witnessing rudeness remained significantly associated with those challenges, though reduced. Additional challenges like witnessing name calling and stereotyping also played significant role in 2024 leading contributors to stop contributing.

USING PSEUDONYM: In 2017, contributors who witnessed impersonation had 6 times higher odds of using a pseudonym compared to those who did not. By 2024, the influence of impersonation had decreased, while name-calling emerged as a strong predictor. Witnessing name-calling in 2024 was associated with a 546% higher odds of using pseudonyms.

WORK PRIVATELY: In 2017, contributors who witnessed rudeness were 10 times more likely to work privately. Witnessing stereotypes also influenced contributors to prefer private work. By 2024, the impact of rudeness had lessened, but name-calling and impersonation became significant factors driving contributors to work privately.

TABLE V: The Impact of Witnessing Interpersonal Challenges on Contributors' Behaviors (2017 and 2024)

	Stop Contributing	Pseudonym	Work Privately	Change Username	Change Online Presence	Suggest COC	Private Discussions	Public Discussions	Offline Changes
Threats of Violence	1.5	-13.07	-56.1	-14.12	-49.24	28.28	-34.8	-2.18	-30.65
Impersonation	42.1	600.8**	22.2	131.02	152.91	182.23**	-2.7	13.81	-1.76
Sustained Harassment	49.1	-82.0	112.3	-50.38	38.71	-11.28	48.0	44.48	185.88
Cross-Platform Harassment	-26.1	140.1	51.6	325.07	2.69	57.02	46.6	28.88	-13.84
Stalking	58.0	38.2	78.0	-21.28	16.53	-10.61	9.33	27.38	169.59
Doxxing	69.7	284.8	15.3	69.34	148.91	-9.28	46.2	-19.96	43.44
Rudeness	2,533.1**	98.43	1,029.8**	76.35	275.45**	687.90**	1775.8**	2041.7 **	383.19**
Name-calling	30.4	40.6	57.2	255.84*	170.60 *	120.39*	56.9*	85.7**	114.36
Stereotyping	53.9	168.5	150.1**	162.67	323.74**	103.33	134.4**	51.46	284.69**
Sexual Advances	-26.3	-4.6	13.5	5.96	-10.76	136.96	40.5	96.49	-41.17
2017 R^2	*0.278*	*0.192*	*0.252*	*0.186*	*0.258*	*0.263*	*0.286*	*0.274*	*0.241*
Threats of Violence	-6.6	40.7	82.9	-1.59	-20.17	17.57	3.0	-11.74	-20.44
Impersonation	48.8	217.7**	106.5**	135.33**	29.11	121.03	28.0	42.57	9.86
Sustained Harassment	42.0	-53.0	65.3	-36.18	122.57*	29.28	56.05	82.32	147.54*
Cross-Platform Harassment	-20.3	18.3	-12.2	-3.16	-19.55	-34.28	46.73	42.70	-17.22
Stalking	45.1	97.3	91.6*	302.74**	35.09	163.04*	54.88	38.50	168.95**
Doxxing	-5.9	72.6	25.2	63.46	86.23	31.91	14.87	56.50	82.21
Rudeness	1,169.2**	10.8	204.9**	23.02	306.84 **	103.51	1244.0**	716.78 **	439.25**
Name-calling	91.8**	546.7**	173.1**	329.88**	124.69**	101.13	80.3**	99.65**	40.64
Stereotyping	94.2**	-16.8	57.3	40.09	58.09	120.39*	44.96	10.14	115.43*
Sexual Advances	48.2	42.7	18.3	65.80	186.06**	123.61	63.29	65.63	63.23
2024 - R^2	*0.267*	*0.189*	*0.257*	*0.217*	*0.250*	*0.222*	*0.281*	*0.253*	*0.265*

* denotes $p < 0.05$ (Bonferroni-corrected, statistically significant), ** denotes $p <= 0.01$ (highly statistically significant).
The light gray columns indicate consequences for which the impact of witnessing interpersonal challenges significantly changed between 2017 and 2024.

SUGGESTING A CODE OF CONDUCT: In 2017, contributors who witnessed rudeness had 6.8 times higher odds of suggesting a code of conduct than those who did not witness it. Impersonation also significantly influenced contributors to advocate for a code of conduct. By 2024, stalking and stereotyping became significant factors.

Overall, the findings reveal a clear shift in the dynamics of contributor behavior from 2017 to 2024. While the influence of rudeness has generally declined over time particularly in its effects on working privately, and stop contributing, other harmful behaviors such as name-calling, impersonation, stereotyping, and sexual advances have remained persistent and significant predictors of negative outcomes. In particular, name-calling and impersonation emerged more prominently in 2024, influencing both private work patterns and the use of pseudonyms. These findings underscore the evolving nature of online interactions and highlight the importance of addressing specific forms of misconduct to foster inclusive and safe contribution spaces.

RQ2 The impact of experiencing and witnessing interpersonal challenges on contributor behavior shifted significantly between 2017 and 2024. In 2017, experiencing rudeness was strongly associated with contributors stopping their participation, working privately, and avoiding offline events. By 2024, both experiencing rudeness and name-calling were strongly associated with these same withdrawal behaviors. Witnessing rudeness remained the strongest predictor in both 2017 and 2024, significantly influencing contributors' decisions to stop contributing and shift towards working privately.

V. DISCUSSION

It is more than time to change the strategies that are not effective. Our findings from RQ1 (Section IV-A) reveal that interpersonal challenges in OSS communities have not only persisted but worsened from 2017 to 2024. This growing toxicity indicates that current measures such as codes of conduct [9] and moderation practices are necessary but insufficient. To foster sustainable contributor engagement, OSS communities must rethink and strengthen their strategies. First, preventive actions must be embedded into project workflows. Communities should proactively monitor healthy communication and provide guidelines, training, and mentorship, rather than reacting only after toxic incidents occur. Second, enforcement mechanisms must be more transparent and consistent, offering visible assurances that reports of misconduct will be taken seriously and addressed fairly. Establishing independent ombudspersons or rotating moderation teams could help. Third, OSS projects should invest in tools that detect and mitigate toxicity early, such as tone detectors or systems encouraging rewording of harsh comments [1, 31]. Finally, leaders must confront systemic inequalities that make certain groups more vulnerable to toxic behavior. Diversifying leadership, ensuring equitable participation, and recognizing positive behaviors are essential to shifting community norms. Without sustained, systemic interventions, OSS communities risk continued attrition and threaten the long-term health and sustainability of their projects.

Rudeness is more pervasive than we may think. As detailed in Section IV-A, rudeness[12] remains the most commonly reported interpersonal challenge, with most contributors experiencing or witnessing (Fig.2) it (Fig. 1) in both 2017 and 2024. It was also the most impactful interpersonal challenge

in 2024 (see Table IV): contributors who experienced rudeness had 46 times higher odds of stopping contributions, 6.7 times higher odds of adopting pseudonyms (potentially undermining contributor identity and trust), 11 times higher odds of working privately or altering their online presence (both reducing transparency and weakening collaboration), and 12 times higher odds of making offline changes such as avoiding in-person conferences (limiting opportunities for community building and networking) than those who did not experience rudeness.

There is still hope in codes of conduct. Although the impact of experiencing interpersonal challenges on suggesting a code of conduct remained relatively stable between 2017 and 2024 (see Table II), certain patterns persisted. Experiencing rudeness continued to be significantly associated with contributors advocating for the adoption of a code of conduct in both years (see Table IV). However, a notable shift occurred in the role of witnessing interpersonal challenges over time. In 2024, witnessing stalking emerged as a new and significant predictor for suggesting a code of conduct an association not observed in 2017 (see Table IV). This rise highlights how stalking, as a form of interpersonal harm, has become increasingly visible and concerning within OSS communities, prompting greater demands for formal governance mechanisms [13]. The emergence of stalking as a trigger for advocacy suggests that contributors are becoming more sensitive to severe boundary violations and are recognizing the need for stronger protective measures to maintain a safe and inclusive environment.

The impact of toxicity in attrition. While decisions to stop contributing to open source projects may stem from a variety of factors, including job demands and personal circumstances unrelated to interpersonal experiences, our 2024 regression model (Table IV) demonstrated a weak-moderate explanatory power ($R^2 = 0.410$). This indicates that experiences with interpersonal challenges account for 41% of the variance in contributors' decisions to leave. Within similar studies in the Software Engineering domain, R^2 values around 0.40 have been considered as weak-moderate predictive power [32]. These findings reinforce prior research showing that interpersonal challenges can undermine contributors' motivation and diminish their willingness to continue participating in OSS [3]. Our results highlighted that toxic interactions particularly rudeness, stalking, name calling, and stereotyping play a substantial role in contributor attrition. OSS communities aiming to retain contributors must take proactive measures to foster more respectful interactions. Strategies such as monitoring code review comments and encouraging constructive, non-rude feedback may help mitigate the negative impact of toxicity and support a healthier, more inclusive community environment.

Some sands don't shift like others. While many behavioral consequences intensified between 2017 and 2024, not all responses to interpersonal challenges shifted significantly over time, as shown in Section IV-B (Tables IV and V). Notably, CHANGING USERNAMES remained a rare and relatively stable behavior. Altering a username disrupts the continuity of a contributor's established reputation, contribution history, and project visibility, making it a costly and often impractical response. Unlike adopting a pseudonym for future contributions or shifting collaboration to private channels, changing a username severs important social and technical ties that are foundational to long-term participation in OSS. This suggests that contributors tend to prioritize strategies that allow them to manage risk while preserving their professional identity and credibility within the community. Even as interpersonal challenges became more frequent and impactful, contributors appeared reluctant to take irreversible actions that might undermine their accumulated social capital. Thus, while the rising tide of toxicity prompted increases in private work, pseudonym adoption, and withdrawal from public spaces, more drastic identity changes remained an uncommon and stable response across the years.

VI. Implications

A. Implications to practice.

Our findings highlight that simply adopting basic measures like codes of conduct [9] and reactive moderation is no longer sufficient to sustain healthy OSS communities. As toxicity has worsened from 2017 to 2024, communities must shift toward proactive, embedded practices that address harmful behaviors before they escalate. Preventive strategies such as integrating communication training, mentorship programs, and inclusive onboarding processes into everyday project workflows are essential. Governance practices must also evolve creating visible, transparent, and community-driven enforcement systems where contributors feel safe reporting interpersonal harm. Given the distinct effects associated with different types of challenges (e.g., stalking vs. rudeness vs. name-calling), interventions must be tailored to address the specific nature and severity of each behavior rather than applying generic responses.

Moreover, maintaining contributor engagement now demands a broader view of contributor safety that extends beyond online platforms. Since toxicity increasingly drives contributors to withdraw from in-person events, OSS communities and event organizers must adopt explicit safety measures for offline interactions as well. Community leaders should actively monitor communication health through regular surveys, community feedback mechanisms, and automated toxicity detection tools. Recognizing and publicly rewarding positive behaviors can also help shift cultural norms toward respect and inclusion. Without these systemic changes, OSS communities risk accelerating attrition, diminishing collaboration, and losing the diverse perspectives critical to long-term project sustainability.

B. Implications to research.

While our study sheds light on the evolving nature of interpersonal challenges in OSS, it also reveals several gaps that future research must address. First, demographic-specific vulnerabilities remain largely unexplored due to dataset limitations. Future studies should investigate how factors such as gender, race, country of origin, immigration status, and language proficiency shape contributors' exposure to toxicity and their behavioral responses. Additionally, it presents an opportunity for investigating how interpersonal challenges

happen across projects having different types of Codes of Conduct (COC), which could enable more targeted and effective interventions.

VII. Threats to Validity

We discuss the threats to validity based on the classification by Wohlin et al. [33]: *construct*, *internal*, *external*, and *conclusion* validity. For each category, we describe the key threats and our mitigation strategies.

A. Construct Validity

Construct validity refers to whether the survey accurately captures contributors' experiences and observations of interpersonal challenges, along with their behavioral consequences. As the study relies on self reported perceptions of incidents like "rudeness," "harassment," and "stalking," interpretations may vary due to personal thresholds, cultural background, or evolving norms. To reduce this variability, the survey used standardized definitions from prior OSS and online harassment research when designing questions. The survey also focused only on challenges consistently measured in both the 2017 and 2024 surveys to ensure comparability. Additionally, the survey distinguished between experiencing and witnessing challenges, improving clarity and reducing ambiguity.

B. Internal Validity

Internal Validity refers to the extent to which observed relationships between interpersonal challenges and behavioral consequences can be attributed to the challenges themselves rather than confounding factors. Because the data are observational and based on retrospective self-reporting, causal inference is limited. External events, such as increased societal awareness of online harassment, or changes in OSS community practices between 2017 and 2024, may have influenced both the prevalence of reported challenges and contributors' behavioral responses. Survivorship bias presents an additional concern, as contributors most severely impacted by early challenges may have exited OSS communities and thus are underrepresented in 2024.

C. External Validity

External validity refers to the generalizability of our findings beyond the studied sample. Since our analysis focuses exclusively on GitHub contributors, applicability to other OSS platforms like GitLab, Bitbucket, or self hosted communities may be limited. Regional, cultural, and linguistic differences can also influence how interpersonal challenges are perceived and reported. Nevertheless, we used large scale GitHub survey data from 2017 and 2024, with thousands of contributors worldwide and surveys available in multiple languages enhancing sample diversity and accessibility. However, the lack of demographic data limits our ability to assess representativeness or determine whether certain groups (e.g., by gender, race, or OSS tenure) were disproportionately affected or responded differently to toxicity. These limitations are further complicated by possible shifts in how contributors define toxic behavior; heightened awareness in 2024 may have led to reporting of issues that were normalized in 2017, potentially influencing trends independently of actual behavior change. Future replication across other platforms and communities is essential to verify and extend our results.

D. Conclusion Validity

Conclusion Validity relates to the reliability and robustness of the statistical conclusions drawn. We employed appropriate statistical techniques including chi-square tests [34], logistic regression [35] models, and likelihood ratio tests [36] to identify significant differences and shifts between 2017 and 2024. However, because the study involves multiple comparisons and exploratory analyses, there remains a risk of Type I errors (false positives) despite using corrections like Bonferroni adjustments [37]. Furthermore, the cross-sectional nature of the surveys limits our ability to trace individual trajectories over time, reducing the granularity of longitudinal conclusions.

VIII. Conclusion

This study analyzed data from two large-scale GitHub Open Source Surveys conducted in 2017 and 2024 to investigate how interpersonal challenges in OSS communities affect contributor behaviors. We examined the prevalence of challenges such as rudeness, name-calling, and stalking, and assessed their association with decisions to disengage, conceal identity, work privately, and withdraw from offline events. Our results show a troubling escalation in both the frequency of these challenges and their negative impacts on participation.

While existing strategies to reduce toxicity have been implemented in OSS and remain important, our findings indicate that they are not sufficient on their own. OSS communities must move beyond reactive measures and invest in proactive, systemic interventions that promote respectful communication, visible enforcement, and cultural change.

Sustaining OSS communities requires ongoing attention to social as well as technical infrastructure. Future work should focus on developing adaptive, context-aware strategies to foster collaboration, inclusion, and resilience in increasingly complex and diverse OSS ecosystems.

Ultimately, as OSS continues to grow in importance for the global digital infrastructure, addressing interpersonal challenges must become a core pillar of project sustainability efforts. The health of OSS ecosystems hinges not only on innovation and technical contributions, but also on the ability to build environments where contributors feel respected, safe, and empowered to collaborate. Recognizing and mitigating toxicity is not just a matter of community well being it is vital to ensuring the long term vitality and success of OSS.

IX. Acknowledgements

We thank the respondents and the GitHub.com authors of the survey: Kenyatta Forbes, Kevin Xu, Jeffrey Luszcz, Margaret Tucker, Eva Maxfield Brown, Peter Cihon, Mike Linksvayer, Ashley Wolf, Lukas Spieß, Kevin Crosby, and Jason Meridth for making the dataset available for research.

References

[1] J. Sarker, "Identification and mitigation of toxic communications among open source software developers," in *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 2022, pp. 1–5.

[2] M. Guizani, A. Chatterjee, B. Trinkenreich, M. May, G. Noa, L. Russell, G. Zambrano, D. Izquierdo-Cortazar, I. Steimacher, M. Gerosa, and A. Sarma, "The long road ahead: Ongoing challenges in contributing to large oss organizations and what to do," *Computer Supported Cooperative Work*, 2021.

[3] M.-A. Storey, A. Zagalsky, F. Figueira Filho, L. Singer, and D. M. German, "How social and communication channels shape and challenge a participatory culture in software development," *IEEE Transactions on Software Engineering*, vol. 43, no. 2, pp. 185–204, 2016.

[4] C. Miller, S. Cohen, D. Klug, B. Vasilescu, and C. Kästner, "" did you miss my comment or what?" understanding toxicity in open source discussions," in *ICSE 2022*, 2022, pp. 710–722.

[5] I. Steinmacher, T. Conte, M. A. Gerosa, and D. Redmiles, "Social barriers faced by newcomers placing their first contribution in open source software projects," in *18th ACM conference on Computer supported cooperative work & social computing*, 2015, pp. 1379–1392.

[6] I. Steinmacher, T. Conte, and M. A. Gerosa, "Understanding and supporting the choice of an appropriate task to start with in open source software communities," in *HICSS'15*. IEEE, 2015, pp. 5299–5308.

[7] C. Jensen, S. King, and V. Kuechler, "Joining free/open source software communities: An analysis of newbies' first interactions on project mailing lists," in *HICSS 2011*. IEEE, 2011, pp. 1–10.

[8] C. Hannebauer and V. Gruhn, "On the relationship between newcomer motivations and contribution barriers in open source projects," in *International Symposium on Open Collaboration*, 2017, pp. 1–10.

[9] P. Tourani, B. Adams, and A. Serebrenik, "Code of conduct in open source projects," in *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. Klagenfurt, Austria: IEEE, 2017, pp. 24–33. [Online]. Available: https://doi.org/10.1109/SANER.2017.7884606

[10] H. Frluckaj and J. Howison, "Codes of conduct in open source," in *Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights*. Apress Berkeley, CA, 2024, pp. 295–308.

[11] M. Lindsay, J. M. Booth, J. T. Messing, and J. Thaller, "Experiences of online harassment among emerging adults: Emotional reactions and the mediating role of fear," *Journal of interpersonal violence*, vol. 31, no. 19, pp. 3174–3195, 2016.

[12] I. Ferreira, J. Cheng, and B. Adams, "The" shut the f** k up" phenomenon: Characterizing incivility in open source code review discussions," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW2, pp. 1–35, 2021.

[13] B. Trinkenreich, Z. Feng, R. Choudhuri, M. Gerosa, A. Sarma, and I. Steinmacher, "Investigating the impact of interpersonal challenges on feeling welcome in oss," *Proceedings of the ACM/IEEE 47th International Conference on Software Engineering*, 2025.

[14] R. Ehsani, M. M. Imran, R. Zita, K. Damevski, and P. Chatterjee, "Incivility in open source projects: A comprehensive annotated dataset of locked github issue threads," in *Proceedings of the 21st International Conference on Mining Software Repositories*, 2024, pp. 515–519.

[15] F. Zlotnick, "GitHub Open Source Survey 2017," http://opensourcesurvey.org/2017/, Jun. 2017.

[16] GitHub, Inc., K. Forbes, K. Xu, J. Luszcz, M. Tucker, E. M. Brown, P. Cihon, M. Linksvayer, A. Wolf, L. Speiß, K. Crosby, and J. Meridth, "GitHub Open Source Survey 2024," Oct. 2024. [Online]. Available: https://github.com/github/open-source-survey

[17] S. Bharadwaj, F. Santos, and B. Trinkenreich, "Replication package for mutation testing comparison," https://figshare.com/s/8c4497eefb7cf07246b0, 2025, available at Figshare.

[18] K.-J. Stol and B. Fitzgerald, "The abc of software engineering research," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 27, no. 3, p. 11, 2018.

[19] M.-A. Storey, N. A. Ernst, C. Williams, and E. Kalliamvakou, "The who, what, how of software engineering research: a socio-technical framework," *Empirical Software Engineering*, vol. 25, pp. 4097–4129, 2020.

[20] K. Finley, "Diversity in open source is even worse than in tech overall," https://www.wired.com/2017/06/diversity-open-source-even-worse-tech-overall/, 2023, accessed: 2024-03-19.

[21] W. E. Powell, D. S. Hunsinger, and B. D. Medlin, "Gender differences within the open source community: An exploratory study," *Journal of Information Technology*, vol. 21, no. 4, pp. 29–37, 2010.

[22] B. Vasilescu, V. Filkov, and A. Serebrenik, "Perceptions of diversity on git hub: A user survey," in *CHASE 2015*. IEEE, 2015, pp. 50–56.

[23] V. Singh and W. Brandon, "Discrimination, misogyny and harassment: Examples from oss: content analysis of women-focused online discussion forums," in *Proceedings of the Third Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering*, 2022, pp. 71–79.

[24] V. Singh, B. Bongiovanni, and W. Brandon, "Codes of conduct in open source software—for warm and fuzzy feelings or equality in community?" *Software Quality Journal*, pp. 1–40, 2021.

[25] R. Ehsani, M. M. Imran, R. Zita, K. Damevski, and P. Chatterjee, "Incivility in open source projects: A comprehensive annotated dataset of locked github issue threads," in *Proceedings of the 21st International Conference on Mining Software Repositories*, ser. MSR '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 515–519. [Online]. Available: https://doi.org/10.1145/3643991.3644887

[26] A. Algarni, Y. Xu, and T. Chan, "Social engineering in social networking sites: The art of impersonation," in *2014 IEEE International Conference on Services Computing*, 2014, pp. 797–804.

[27] M. L. Pittaro, "Cyber stalking: An analysis of online harassment and intimidation," *International journal of cyber criminology*, vol. 1, no. 2, pp. 180–197, 2007.

[28] P. Q. Brady and M. R. Nobles, "The dark figure of stalking: Examining law enforcement response," *Journal of interpersonal violence*, vol. 32, no. 20, pp. 3149–3173, 2017.

[29] N. Dragan, M. L. Collard, and J. I. Maletic, "Automatic identification of class stereotypes," in *2010 IEEE International Conference on Software Maintenance*, 2010, pp. 1–10.

[30] S. Dubberley, A. Koenig, and D. Murray, *Digital witness: using open source information for human rights investigation, documentation, and accountability*. Oxford University Press, USA, 2020.

[31] R. Ehsani, R. Rezapour, and P. Chatterjee, "Exploring moral principles exhibited in oss: A case study on github heated issues," in *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2023, pp. 2092–2096.

[32] B. Trinkenreich, 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 *ICSE 2023*. IEEE, 2023.

[33] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer Nature, 2024.

[34] T. M. Franke, T. Ho, and C. A. Christie, "The chi-square test: Often used and more often misinterpreted," *American Journal of Evaluation*, vol. 33, no. 3, pp. 448–458, 2011, original work published 2012. [Online]. Available: https://doi.org/10.1177/1098214011426594

[35] T. M. Khoshgoftaar and E. B. Allen, "Logistic regression modeling of software quality," *International Journal of Reliability, Quality and Safety Engineering*, vol. 6, no. 4, pp. 303–317, 1999. [Online]. Available: https://doi.org/10.1142/S0218539399000292

[36] J. T. Kent, "Robust properties of likelihood ratio tests," *Biometrika*, vol. 69, no. 1, pp. 19–27, apr 1982. [Online]. Available: https://doi.org/10.1093/biomet/69.1.19

[37] J. F. Morgan, "p value fetishism and use of the bonferroni adjustment," *Evidence Based Mental Health*, vol. 10, no. 2, pp. 34–35, 2007. [Online]. Available: https://ebmh.bmj.com/content/10/2/34