Exploring Moral Principles Exhibited in OSS: A Case Study on GitHub Heated Issues

Ramtin Ehsani
Drexel University
Philadelphia, PA, USA
ramtin.ehsani@drexel.edu

Rezvaneh Rezapour
Drexel University
Philadelphia, PA, USA
shadi.rezapour@drexel.edu

Preetha Chatterjee
Drexel University
Philadelphia, PA, USA
preetha.chatterjee@drexel.edu

ABSTRACT

To foster collaboration and inclusivity in Open Source Software (OSS) projects, it is crucial to understand and detect patterns of toxic language that may drive contributors away, especially those from underrepresented communities. Although machine learning-based toxicity detection tools trained on domain-specific data have shown promise, their design lacks an understanding of the unique nature and triggers of toxicity in OSS discussions, highlighting the need for further investigation. In this study, we employ Moral Foundations Theory to examine the relationship between moral principles and toxicity in OSS. Specifically, we analyze toxic communications in GitHub issue threads to identify and understand five types of moral principles exhibited in text, and explore their potential association with toxic behavior. Our preliminary findings suggest a possible link between moral principles and toxic comments in OSS communications, with each moral principle associated with at least one type of toxicity. The potential of MFT in toxicity detection warrants further investigation.

CCS CONCEPTS

• **Software and its engineering** → *Software organization and properties.*

KEYWORDS

moral principles, toxicity, open source, textual analysis

1 INTRODUCTION

Open source software (OSS) projects have become crucial for the digital advancement of society, with over 90% of companies leveraging open source [1]. However, many new and longstanding OSS projects fail, in part due to toxic interactions and uncivil language,

which serve as a barrier to contribution and lead to negative emotions and isolation [13, 27, 37]. Thus, understanding toxicity in OSS has gained increasing attention in recent years [7–9, 11, 34, 35]. Uncivil features such as name-calling, frustration, and impatience were found in more than half of non-technical emails in the Linux Kernel mailing list [10], swearing and personal offense were identified as the main reasons for offense in several communities [7], and GitHub issue threads contain various types of toxic behaviors [23]. Although researchers outside the Software Engineering (SE) domain have designed several techniques to automatically detect toxicity in online forums [22, 25, 36], directly applying those tools to SE is ineffective due to several factors, including the unique nature of SE text [18, 29, 34]. More recently, machine learning-based tools have shown promise in detecting toxicity when trained on domain-specific data [35], however it remains crucial to address the fundamental issue of their design lacking an understanding of the unique nature and triggers of toxicity in OSS [23, 28].

Language is an indispensable tool to convey information between people, and analyzing it could play a significant role in understanding the challenges faced in OSS communities [12, 32, 33]. Linguistic styles and word choices can help in analyzing and understanding people's thoughts, opinions, and feelings [3, 6, 31, 39]. People's values and personal norms affect their (spontaneous) attitude, decision-making process, and what they perceive as good or bad, and moral or immoral [15, 30]. Previous research in psychology has addressed the connection between moral principles and hate and showed that morality is a key feature of hatred; hate is connected to core moral beliefs and higher levels of moral emotions (e.g., contempt, anger, disgust) [2, 26]. Considering the importance of effective communication, we believe that analyzing moral values as exhibited in OSS texts can provide valuable insights into the underlying beliefs and values that influence communication styles in this community, especially in spreading toxicity.

In this paper, we aim to investigate the moral principles exhibited in GitHub and explore their relationship with toxicity. To that end, we leverage Moral Foundation Theory (MFT), a social psychological theory that assumes individual judgments are influenced by emotional and cognitive appraisals, referred to as intuitions or foundations [14, 16]. We explore a dataset of 100 toxic issue threads labeled with their representative natures (e.g., 'Insulting', 'Arrogant') [23], to analyze moral principles and how they are exhibited in the context of OSS project communications. We map the observed moralities in each thread to their corresponding natures of toxicity to understand the relationship and co-occurrence. Our findings reveal that toxic issue threads exhibit moral principles, with each moral principle being associated with at least one type of toxicity. This suggests that toxic comments are associated with moral principles, and that the Moral Foundations Theory (MFT)

may be useful in detecting toxicity. To apply MFT to the domain of SE, it would be necessary to adapt it, and we believe that doing so has the potential to enhance our understanding and detection of toxicity in SE communications.

2 MORAL FOUNDATIONS THEORY

With the goal of investigating instances on GitHub that represent moral principles exhibited in the text, we leverage Moral Foundations Theory (MFT) [14]. MFT classifies human behavior into five fundamental principles that represent contrasting values:

Care/harm: This principle is based on our general dislike of suffering, whether for ourselves or others, related to our evolution as mammals. It underlies the virtues of kindness, compassion, and gentleness, and it condemns cruelty and aggression.

Fairness/cheating: This principle relates to justice and rights, and is linked to the evolutionary process of reciprocal altruism.

Loyalty/betrayal: This principle promotes patriotism, heroism, trust, and self-sacrifice for the group, rooted in our history as tribal beings. It values the virtue of "One for all, and all for one," while considering acts of betrayal towards social structures as immoral.

Authority/subversion: This principle is shaped by our primate history of hierarchical social structure, resulting in virtues such as leadership, followership, and deference to authority/traditions. It may view dissent against authority as immoral.

Sanctity/degradation: This principle arises from the psychology of disgust and contamination, where the concept of humans striving to live in a noble and elevated way is the key. It reflects the idea of the body as a temple that can be corrupted by immorality.

3 METHODOLOGY

To understand how moral principles manifest in OSS project's textual communication and its relation to toxicity, we analyze the GitHub developer interactions. Our focus is on GitHub heated issues, which often result from conflicts, as they can provide valuable insights into the ethical and moral considerations of participants. It should be noted that we are only analyzing textual representations of moral values, and our analysis does not represent the moral values of people engaging in these discussions.

Dataset. We analyze a benchmark dataset of 100 toxic GitHub issue comment threads curated by Miller et al. [23], in which toxicity is defined as "rude, disrespectful, or unreasonable language that is likely to make someone leave a discussion". The dataset includes information on the issue link, author, trigger, target, and the nature of toxicity for each thread of comments. Out of the 100 issue threads, 20 contain comments that were either removed or the project along with its issues were taken down from GitHub. The 5 categories of toxic instances in this dataset are: *Insulting*, *Entitled*, *Arrogant*, *Trolling*, and *Unprofessional*.

Procedure. To understand the characteristics and patterns of morality in the context of SE, we qualitatively analyze 100 issue threads and their context, using thematic analysis [4], and following the trustworthiness criteria (credibility, transferability, dependability, and confirmability) as recommended by previous research [21, 24]. The data analysis process was conducted iteratively and reflectively, following recommendations from qualitative analysis [20]. To enhance our understanding of the developers' interactions and the

contextual factors influencing their communication, we first carefully read through all the issue comment threads. Next, the goal was to associate moral values (as defined in Section 2) to each comment in the threads, if applicable. Given the substantial differences between OSS communications and other social media platforms such as Twitter and Reddit, we adapted the original definitions of morality to align them with the context of software engineering. Specifically, we identify key concepts of each morality type, observe patterns in the dataset, and use them to create the mapping of morality categories. Throughout the analysis, careful notes were taken to capture patterns of morality observed in the comments. This process was repeated multiple times, with constant additions to the notes and observations, and ongoing reevaluation of the mappings in each iteration.

Note that in the process of labeling morality to the issue comments, we consciously disregard the toxicity labels to mitigate bias in our decisions, so that we do not associate a specific type of toxicity with a certain type of morality. The result of our mapping between the morality categories and their manifestation in GitHub heated issues is presented in Table 1.

4 PRELIMINARY OBSERVATIONS

In this section, we discuss our observations for the moral principles as exhibited in the GitHub heated issues. We analyze the issue comment threads in our data sample to identify instances where moral principles were expressed in the text. Out of 695 issue comments across 100 threads, 135 exhibit at least one form of moral principle. In this analysis, people who open the issue threads to find fixes for problems are referred to as *users*, and people who try to resolve the issues and close the threads as *contributors*. We observed unique patterns specific to each of the identified roles.

Care/Harm. A total of 48 comments exhibit this principle. Pertaining to this category, following are the behaviors we observed: **(CH1)** users who use derogatory language or insults towards contributors while seeking assistance or requesting new features; **(CH2)** users insulting contributors mainly by trolling, possibly due to differences in ideology, opinions, etc.; **(CH3)** contributors responding using insults toward users.

The first type (CH1) was the most common in our dataset (32 out of 48 instances). For example, in one thread, a user opens an issue saying that the name of the project is duplicated and it should change while using insults: *"This name is in use and should be changed immediately. Yes, I have seen the other issues. Yes, I am opening a new one because f**k you Microsoft. You are merely trying to cast a shadow on other, truly open source projects...".*

The second type (CH2) was also seen in a number of cases (5 out of 48 instances). For example, in one instance, a user opens an issue saying: *"Revenue. F**k you guys"*, and does not add anything else to the thread. Threads similar to this are mostly the result of trolling, but they can also be the result of differences in ideology or based on previous interactions [23].

For the third type (CH3), we saw an interesting pattern of contributors responding in a detrimental manner towards the users. This type is rare compared to other types in this category. In one instance, a user opens an issue asking: *"Where to f**k python2? Why, when I give the brew install python command, python3 is installed, not python2. Developers are you stoned there?"*, and surprisingly, one of

Moral Principle — Def.	As Exhibited in Issue Threads
Care/harm — Protecting versus hurting others	**Care:** Developer communications exhibiting kindness, e.g., experienced contributors shielding newcomers from criticism. **Harm:** Developers exchanging insults in issue threads, either in response to feature requests or bug fixes, or due to trolling or experience from past interactions [23]).
Fairness/cheating — Cooperation/ trust/ just versus cheating in interaction with objects and people	**Fairness:** Developers and contributors respecting one another's rights (e.g., having the right to open threads and ask for features/fixes, get the help they need). **Cheating:** Disregarding developers' rights by imposing unrealistic expectations (e.g., unfeasible project timeline), or by discriminating against them (e.g., based on gender) and not addressing their concerns. [41].
Loyalty/betrayal — Ingroup commitment (to coalitions, teams, brands) versus leaving group	**Loyalty:** Acts of self-sacrifice and altruism for the community, ranging from dedicating time to participating in issues threads to defending developers and the project itself against criticisms. **Betrayal:** OSS communities, like all social communities, view acts of betrayal as a breach of community morale. Betrayal is observed when developers urge others to abandon a project, or exclude them from the community.
Authority/subversion — Adhering to the rules of hierarchy versus challenging hierarchies	**Authority:** OSS communities often have guidelines that must be followed, e.g., Code of Conduct: Authorities enforcing CoC or any admin-privileged acts (minimizing comments, closing issues, etc.) to ban or censor the users. **Subversion:** Developers trying to rebel against authority, and questioning contributors' ability to lead the community.
Sanctity/degradation — Behavioral immune system versus spontaneous reaction	**Sanctity:** Communicating with temperance and respect for one another, and encouraging others to do so as well. **Degradation:** Developers expressing their hatred (disgust) toward a certain package, system, or code in the project. This behavior is targeted at code (objects) rather than people.

Table 1: Moral Principles in GitHub Heated Issues

the contributors responds with: *"I recommend not having intercourse with EOL software. Python 2 is no longer included in Homebrew"*. The user then responds: *"You have an unfinished raw product with some problems. What can be your attitude? Treating users like sh*t"*. This issue thread exhibits the first and last patterns (CH1 and CH3).

Fairness/Cheating. We observed behaviors associated with this category in a total of 19 comments, which include: **(FC1)** users having unrealistic expectations from the contributors; **(FC2)** contributors failing to address user's issues due to unjustified reasons.

Among the instances of this category, 7 out of 19 instances exhibit the first type of communication pattern. In one of the instances representing the first type (FC1), the user has opened an issue asking about a problem in the debugger. Contributors of the projects try to help the user by providing tips on how the problem can be solved, but as the conversation goes on, the user gets frustrated and says: *"@<USER1> @<USER2> Too busy to respond?"*, which in the end, leads to the issue thread being locked. This is an example of unrealistic expectations about response time causing conflict.

For the second type (FC2), for instance, this is how one of the contributors responds to a user's issue: *"How would be this a priority, the app must be ready to be shipped in a few days and basically, no one uses the smartphone in landscape mode except for watching media"*.

Another type of behavior related to *Fairness* would be discrimination (based on race, culture, gender, religion, etc). In our dataset, we did not observe any instances of discrimination. Analysis of additional data may reveal discriminatory behavior in OSS, as such occurrences have been observed in previous studies [7, 41].

Loyalty/Betrayal. The observed patterns for this category of morality were found in 23 instances of the dataset: **(LB1)** users actively promoting and inciting rebellion against a project, encouraging others to switch to an alternative project (most common with 15 instances); **(LB2:)** contributors excluding users from the project.

To better understand these behaviors, we provide examples for each type. For instance, in one of the issue threads, a user claims that ssl insecure is not being respected in the project: *"Whoever will find this issue and gets pissed off, because the author doesn't bother to fix it for years, ditch mitmproxy and use SSL SPLIT"*. One of the developers responds with: *"I'm glad you have found something that works for you!"*. In this instance, both LB1 and LB2 were observed, because the user is encouraging people to leave this project behind, and the contributor also encourages the user to not use the project. In another instance, a user requests the removal of the slur filter in the app, however, one of the contributors says: *"If you don't like it, fork it. Stop bothering us about it, we will never fully remove the slur filter"*, which counts as excluding the user from the community.

Authority/Subversion. OSS communities, like any other community, often have specific rules and guidelines that everyone has to follow. GitHub projects use the Code of Conduct (CoC) in an attempt to promote their expectations and standards of ethical behavior within the community [38]. The establishment of such rules and hierarchies often leads to the dynamic interplay of moralities such as authority and subversion in OSS communications. A total of 30 instances with this principle were found in our dataset. The patterns observed for this category are: **(AS1)** users trying to rebel against authority, and questioning developers' ability to lead the community. **(AS2)** contributors (authorities) enforcing Code of Conduct when thought necessary or any other admin-privileged acts to ban, censor, or silence users (most common with 16 instances).

In one of the instances, a user has opened an issue, claiming that the project has stolen his code: *"Don't be a d*ck as the source code was stolen at that point and implemented on this"*. The contributor enforces authority by saying: *"please be civilized and refrain from profanities as required by our Code of Conduct, or I'll ban you from all Falconry projects. This is the last warning"*. And the user challenges the authority by saying: *"I'm not a bot but if you wish to ignore me go ahead and I'll get a lawyer"*. Additionally, there were instances where the users questioned the capabilities of contributors. For example, a user opened an issue asking about how to do a certain thing within the system. As the conversation goes on, the user responds to one of the contributors: *"...I would hate to have you in charge of any security issues"*.

Sanctity/Degradation. The most commonly observed pattern in this category, in a total of 42 instances, was observed when: **(SD1)** users express their hate (disgust) toward a certain package, system, or code that the project is using, most likely due to their personal
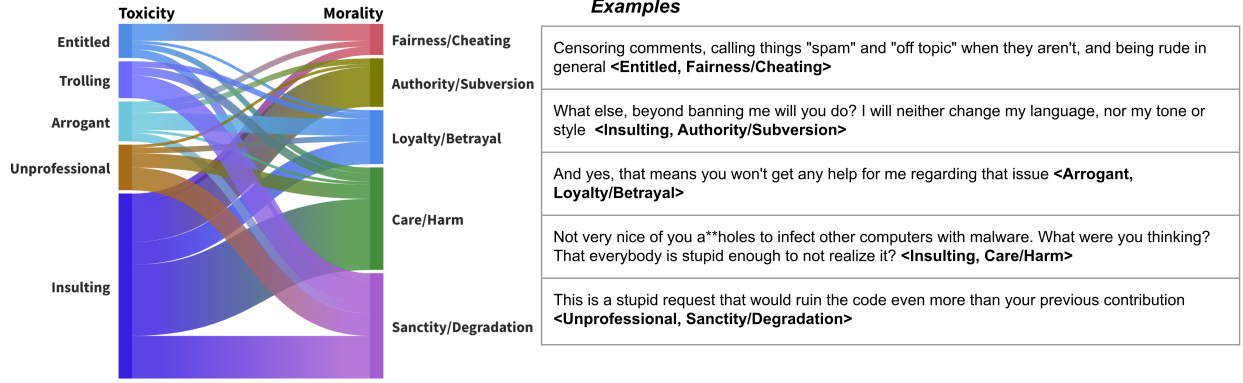
Figure 1: Morality and Toxicity: Examples of their Co-occurrence as Exhibited in GitHub

preferences. This principle is targeted at code/system rather than people (targeting people represents *Care/Harm*).

For example, a user opened an issue in the project expressing how much the document is poorly written, and in a sense, he is *disgusted* with it: *"...in fact, you write a sh*t doc. I'm a real man, it's my feeling of your holy sh*t doc".* Or in another instance, a user opens an issue explaining his problem with the software: *"...I just tried reinstalling you buggy, sh*tty software for the third time...".*

After gathering our observations of the exhibited patterns of morality in GitHub communications, we associated the toxicities presented in our dataset with the moral principles in each issue thread. The results generated for this mapping are shown in Figure 1. Based on this mapping, we observe that every type of toxicity is associated with at least one form of morality. Entitlement is mostly associated with the principle of *Fairness/Cheating*, Trolling with *Sanctity/Degradation*, Arrogant to *Loyalty/Betrayal*, Unprofessional with *Sanctity/Degradation*, and Insulting with *Care/Harm*.

The relationship between the concepts of toxicity and moral principles can be observed in their respective definitions. For example, entitlement comes from disregarding people's rights for personal gain. This aligns with the *Fairness/Cheating* definitions given in previous sections. The relationship between *Authority/Subversion* and insulting threads implies that using insults causes the authorities to intervene more in the issues compared to the other types of toxicity. The moralities most frequently observed in toxic threads are *Sanctity/Degradation* and *Care/Harm*, which were primarily expressed through insults directed towards individuals or code.

5 IMPLICATIONS AND CHALLENGES

We found that toxic behaviors can be categorized and linked to various types of moral principles, indicating the potential for utilizing moral principles to detect toxicity in OSS communications. To this end, numerous Moral Foundations Dictionaries (MFD) can be integrated into models to detect moral values in SE texts, as demonstrated in other fields [19, 31]. Recent studies have explored the association between moral values and the sense of belonging in OSS communities [40]. Adapting existing morality dictionaries to SE and incorporating them into detection tools could facilitate a better understanding of several human values in OSS, including toxicity. It is worth noting that our study only examined the association between moral principles and toxicity in OSS communities, and did not explore any potential causal relationships between the two. While our findings suggest that moral principles are associated

with toxic interactions, it is possible that other factors may also contribute to toxic behavior in OSS communities.

We acknowledge that inferring human values based on textual communication data has its perils. For instance, previous studies show high error rate in detecting developer personalities based on textual data [5, 42]. We emphasize that we are only analyzing the textual representation of moral principles, and we are not associating morality values with the people involved in the discussions. Our findings of the exhibited moralities and their association with toxicity are limited to a relatively small dataset of GitHub issue threads. Further analysis on a large and diverse dataset is necessary to generalize the findings. Detecting morality even in social contexts (e.g., Twitter) is a hard task, mostly in distinguishing co-occurring moral principles [17]. We attempted to address that challenge by mapping and creating clear definitions of moral principles in the domain of SE (Table 1). Another important point to consider is the potential impact of cultural and social differences on the interpretation and expression of moral principles in OSS communities. Our study focused on English texts, and it is possible that different cultures and languages may have different conceptions and expressions of moral principles. Therefore, future research should aim to investigate the role of cultural and linguistic factors in shaping the expression and interpretation of moral principles in OSS communities.

6 CONCLUSION

This paper investigates moral principles and their relation with toxicity in OSS. Due to the use of domain-specific words and jargon, the majority of toxicity detection tools do not correctly identify toxic language in OSS. Therefore, there is a need to augment the current methods with informative features, and to gain a comprehensive understanding of the triggers of toxicity and its underlying causes. In this work, we leverage a toxicity-labeled dataset of GitHub to analyze the key concepts of five moral principles and identify exhibited patterns in the dataset. We also mapped the nature of toxicity to the observed types of moral principles in each thread. Our findings revealed that *Sanctity/Degradation* and *Care/Harm* were the most frequently observed moral principles, and were often associated with insulting threads. Overall, our study represents an initial step towards a deeper understanding of the role of moral principles in shaping community dynamics in OSS, and we hope that it will inspire further research in this important area. Consistent with Miller et al's [23] data distribution policy, we will share our annotated dataset upon request for research purposes.

REFERENCES

[1] 2022-11-10. *GitHub's Octoverse report finds 97% of apps use open source software.* https://venturebeat.com/programming-development/github-releases-open-source-report-octoverse-2022-says-97-of-apps-use-oss/

[2] Mohammad Atari, Aida M Davani, Drew Kogon, Brendan Kennedy, Nripsuta A Saxena, Ian A Anderson, and Morteza Dehghani. 2021. Morally Homogeneous Networks and Radicalism. https://doi.org/10.1177/19485506211059329

[3] Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin* (2022).

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa

[5] Fabio Calefato and Filippo Lanubile. 2022-07-31. Using Personality Detection Tools for Software Engineering Research: How Far Can We Go? 31, 3 (2022-07-31), 1–48.

[6] Preetha Chatterjee, Kostadin Damevski, and Lori Pollock. 2021. Automatic Extraction of Opinion-based Q&A from Online Developer Chats. In *Proceedings of the 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. 1260–1272. https://doi.org/10.1109/ICSE43902.2021.00115

[7] Jithin Cheriyan, Bastin Tony Roy Savarimuthu, and Stephen Cranefield. 2021-06-21. Towards offensive language detection and reduction in four Software Engineering communities. In *Evaluation and Assessment in Software Engineering* (New York, NY, USA) *(EASE 2021)*. Association for Computing Machinery, 254–259. https://doi.org/10.1145/3463274.3463805

[8] Carolyn D. Egelman, Emerson Murphy-Hill, Elizabeth Kammer, Margaret Morrow Hodges, Collin Green, Ciera Jaspan, and James Lin. 2020-10-01. Predicting developers' negative feelings about code review. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering* (New York, NY, USA) *(ICSE '20)*. Association for Computing Machinery, 174–185.

[9] Isabella Ferreira, Bram Adams, and Jinghui Cheng. 2022-05-23. How heated is it? Understanding GitHub locked issues. In *Proceedings of the 19th International Conference on Mining Software Repositories*. 309–320. arXiv:2204.00155 [cs]

[10] Isabella Ferreira, Jinghui Cheng, and Bram Adams. 2021-10-13. The "Shut the f**k up" Phenomenon: Characterizing Incivility in Open Source Code Review Discussions. 5 (2021-10-13), 1–35. Issue CSCW2. arXiv:2108.09905 [cs]

[11] Isabella Ferreira, Ahlaam Rafiq, and Jinghui Cheng. 2022-07-07. Incivility Detection in Open Source Code Review and Issue Discussions.

[12] Jennifer Ferreira, Michael Glynn, David Hunt, Jaganath Babu, Denis Dennehy, and Kieran Conboy. 2019-08-20. Sentiment analysis of open source communities: an exploratory study. In *Proceedings of the 15th International Symposium on Open Collaboration* (New York, NY, USA) *(OpenSym '19)*. Association for Computing Machinery, 1–5.

[13] Marco Gerosa, Igor Wiese, Bianca Trinkenreich, Georg Link, Gregorio Robles, Christoph Treude, Igor Steinmacher, and Anita Sarma. 2021. The Shifting Sands of Motivation: Revisiting What Drives Contributors in Open Source. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. 1046–1058.

[14] Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. 2013. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*. Vol. 47. Elsevier.

[15] Jonathan Haidt. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review* 108, 4 (2001), 814.

[16] Jonathan Haidt and Craig Joseph. 2004. Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus* 133, 4 (2004), 55–66.

[17] Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, Gabriela Moreno, Christina Park, Tingyee E. Chang, Jenna Chin, Christian Leong, Jun Yen Leung, Arineh Mirinjian, and Morteza Dehghani. 2020-11-01. Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. 11, 8 (2020-11-01), 1057–1071. Publisher: SAGE Publications Inc.

[18] Mia Mohammad Imran, Yashasvi Jain, Preetha Chatterjee, and Kostadin Damevski. 2022. Data Augmentation for Improving Emotion Recognition in Software Engineering Communication. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering (ASE)*.

[19] Brendan Kennedy, Preni Golazizian, Jackson Trager, Mohammad Atari, Joseph Hoover, Aida Mostafazadeh Davani, and Morteza Dehghani. 2022. The (Moral) Language of Hate. https://doi.org/10.31234/osf.io/eqp34

[20] Sarah Lewis. 2015. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. *Health Promotion Practice* 16, 4 (2015), 473–475. https://doi.org/10.1177/1524839915580941 arXiv:https://doi.org/10.1177/1524839915580941

[21] Yvonna S. Lincoln and Egon G. Guba. 1985-04. *Naturalistic Inquiry*. SAGE. Google-Books-ID: 2oA9aWlNeooC.

[22] Suman Kalyan Maity, Aishik Chakraborty, Pawan Goyal, and Animesh Mukherjee. 2018-11-01. Opinion Conflicts: An Effective Route to Detect Incivility in Twitter. 2 (2018-11-01), 117:1–117:27. Issue CSCW. https://doi.org/10.1145/3274386

[23] Courtney Miller, Sophie Cohen, Daniel Klug, Bogdan Vasilescu, and Christian Kästner. 2022. "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In *2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE)*. 710–722. https://doi.org/10.1145/3510003.3510111

[24] Lorelli Nowell, Jill Norris, Deborah White, and Nancy Moules. 2017-10-03. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. 16 (2017-10-03).

[25] Paola Pascual-Ferrá, Neil Alperstein, Daniel J Barnett, and Rajiv N Rimal. 2023-04-26. Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. 8, 1 (2023-04-26), 20539517211023533. https://doi.org/10.1177/20539517211023533 Publisher: SAGE Publications Ltd.

[26] Clara Pretus, Jennifer L Ray, Yael Granot, William A Cunningham, and Jay J Van Bavel. 2018. The psychology of hate: Moral concerns differentiate hate from dislike. https://doi.org/10.31234/osf.io/x9y2p

[27] Huilian Sophie Qiu, Yucen Lily Li, Susmita Padala, Anita Sarma, and Bogdan Vasilescu. 2019. The Signals That Potential Contributors Look for When Choosing Open-Source Projects. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 122 (nov 2019), 29 pages. https://doi.org/10.1145/3359224

[28] Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kästner, Carolyn Egelman, Ciera Jaspan, and Emerson Murphy-Hill. 2022. Detecting Interpersonal Conflict in Issues and Code Review: Cross Pollinating Open- and Closed-Source Approaches. In *2022 IEEE/ACM 44th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. 41–55. https://doi.org/10.1109/ICSE-SEIS55304.2022.9793879

[29] Naveen Raman, Minxuan Cao, Yulia Tsvetkov, Christian Kästner, and Bogdan Vasilescu. 2020. Stress and Burnout in Open Source: Toward Finding, Understanding, and Mitigating Unhealthy Interactions. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results* (Seoul, South Korea) *(ICSE-NIER '20)*. Association for Computing Machinery, New York, NY, USA, 57–60. https://doi.org/10.1145/3377816.3381732

[30] Rezvaneh Rezapour, Ly Dinh, and Jana Diesner. 2021. Incorporating the measurement of moral foundations theory into analyzing stances on controversial topics. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media.*

[31] Rezvaneh Rezapour, Saumil H Shah, and Jana Diesner. 2019. Enhancing the measurement of social effects by capturing morality. In *Proceedings of the tenth workshop on computational approaches to subjectivity, sentiment and social media analysis.*

[32] Amirali Sajadi, Kostadin Damevski, and Preetha Chatterjee. 2023. Interpersonal Trust in OSS: Exploring Dimensions of Trust in GitHub Pull Requests. In *Proceedings of the 45th International Conference on Software Engineering (NIER Track)* (Melbourne, Australia) *(ICSE '23)*.

[33] Arghavan Sanei, Jinghui Cheng, and Bram Adams. 2021-03-18. The Impacts of Sentiments and Tones in Community-Generated Issue Discussions. https://doi.org/10.48550/arXiv.2103.10615 arXiv:2103.10615 [cs]

[34] Jaydeb Sarker, Asif Kamal Turzo, and Amiangshu Bosu. 2020. A Benchmark Study of the Contemporary Toxicity Detectors on Software Engineering Interactions. *2020 27th Asia-Pacific Software Engineering Conference (APSEC)* (2020), 218–227.

[35] Jaydeb Sarker, Asif Kamal Turzo, Ming Dong, and Amiangshu Bosu. 2022. Automated Identification of Toxic Code Reviews Using ToxiCR. https://doi.org/10.48550/ARXIV.2202.13056

[36] Amit Sheth, Valerie L. Shalin, and Ugur Kursuncu. 2021-04-21. Defining and Detecting Toxicity on Social Media: Context and Knowledge are Key. (2021-04-21). arXiv:2104.10788 http://arxiv.org/abs/2104.10788

[37] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. 2015. Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1379–1392. https://doi.org/10.1145/2675133.2675215

[38] Parastou Tourani, Bram Adams, and Alexander Serebrenik. 2017-02. Code of conduct in open source projects. In *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. 24–33. https://doi.org/10.1109/SANER.2017.7884606

[39] Harry C Triandis. 1989. The self and social behavior in differing cultural contexts. *Psychological review* 96, 3 (1989), 506.

[40] Bianca Trinkenreich, Klaas-Jan Stol, Anita Sarma, Daniel M. German, Marco A. Gerosa, and Igor Steinmacher. 2023-02-22. Do I Belong? Modeling Sense of Virtual Community Among Linux Kernel Contributors.

[41] Sterre van Breukelen, Ann Barcomb, Sebastian Baltes, and Alexander Serebrenik. 2023. "STILL AROUND": Experiences and Survival Strategies of Veteran Women Software Developers. In *Proceedings of the 45th International Conference on Software Engineering (ICSE 2023)*. arXiv:2302.03723 [cs.SE]

[42] Frenk van Mil, Ayushi Rastogi, and Andy Zaidman. 2021-07-15. Promises and Perils of Inferring Personality on GitHub. arXiv:2107.05829 [cs]