How People Initiate and Respond to Discussions Around Online Community Norms: A Preliminary Analysis on Meta Stack Overflow Discussions

Jingchao Fang
jcfang@ucdavis.edu
University of California, Davis
Davis, California, USA

Jia-Wei Liang
jwlliang@ucdavis.edu
University of California, Davis
Davis, California, USA

Hao-Chuan Wang
hciwang@ucdavis.edu
University of California, Davis
Davis, California, USA

ABSTRACT

Supporting norms-related discussions can aid people in understanding and abiding by ambiguous norms in large-scale online communities. Yet, how social and linguistic factors, such as the identities of interlocutors and the language framing of posts, can influence discussions around norms, is underexplored. In this work, we performed a preliminary analysis based on a dataset containing 123 question threads on Meta Stack Overflow, a site for discussions of the workings and policies of Stack Overflow, to understand how people initiate and respond to norms-related discussions. Results revealed that question posts with different levels of personal relatedness and question specificity have significantly different sentiments, and they also draw comments with diverged sentiments. We present implications and directions of future work based on our findings.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**.

KEYWORDS

online community; community norms; sentiment analysis

1 INTRODUCTION

The success of an online community relies heavily on people's compliance with the norms that are co-constructed by the community members. However, in large-scale online communities, norms are usually implicit, ambiguous, and constantly evolving. Identifying and following online community norms are especially difficult for novices [3, 4, 12]. Therefore, maintaining a communication channel that enables people to discuss questions around norms becomes important.

In this paper, we took Stack Overflow [1] as an example of online communities and studied how people ask and react to questions about community norms. Stack Overflow, a knowledge-based online programming community, has been recognized as a highly successful model of large-scale CSCW systems in its early years of establishment, evidenced by its large volumes of monthly visits, high answer rate, and short answer time [8]. However, as the community scales up, Stack Overflow seems to be thriving less. Many questions ended up being unanswered [11], and the community was criticized to be unfriendly, especially to novices [12]. Conflicts emerge on Stack Overflow as users do not abide by the norms, while a significant portion of norm violation is due to people's unfamiliarity with the norms. In the interest of providing the community a space for discussing the workings of Stack Overflow and approaching self-governance, Meta Stack Overflow [2] was launched in 2014. Since then, a great number of discussions pertinent to community norms and policies have unfolded on it.

One observation is that posting on Meta Stack Overflow is oftentimes triggered by negative experiences such as ill-received contributions (e.g., posts being downvoted or deleted) and conflicts on Stack Overflow, and discussions under such circumstances can sometimes be hard to navigate as they include different stances from the community. We focused on understanding how such discussions are influenced by various social and linguistic factors. Inspired by patterns that emerged on Stack Overflow and the construal level theory (CLT), a conceptual model that describes the effect of psychological distance on thinking abstractness [13], we analyzed how discussion initiators' experience levels and the language framing of their posts influence the community's reactions and conversational patterns. Analysis results showed that non-personal and non-specific posts that initiate discussions around community norms are generally more well-received and draw positive responses. In continuation of previous CSCW research in online community building (e.g., [2, 6]), our work sheds light on how online community maintainers can design guidance for community members to establish effective and positive norms-related discussions and keep the community welcoming.

[1] https://stackoverflow.com/
[2] https://meta.stackoverflow.com/

2 BACKGROUND AND HYPOTHESES

2.1 Experience Levels Matters in Online Communities

People's interactions within online communities are largely influenced by their experience levels. For example, compared to experienced users, novices' questions are more likely to be ill-received (e.g., unanswered, criticized, or deleted) due to their unfamiliarity with community norms and insufficient knowledge of the platform mechanisms on Stack Overflow [1, 3, 4]. We expect that a similar pattern may also be observed in norms-related discussions on Meta Stack Overflow since (1) The reputation score on Meta Stack Overflow is the same as the reputation score on Stack Overflow for each user (which is defined as a measure of how "trustable" the user is), and (2) The user groups of the two sites overlap. Specifically, we pose H1:

H1. *The experience level of question-askers can affect how people react to the questions about community norms. Questions asked by people who have high reputations are more well-received.*

2.2 Construal Level Theory and Emotion Regulation

Questions about norms can be framed in different ways and may attract varied reactions from the community. **Construal level theory** (CLT) is a conceptual framework stating that more psychologically distant objects or events are construed at a higher level, and thus lead to more abstract mental processing [13]. In this paper, we specifically consider **personal relatedness** and **question specificity** of Meta Stack Overflow posts as two factors that can reflect psychological distance, and investigate their effects on the sentiment in language use.

2.2.1 Personal Relatedness. Some questions about community norms on Meta Stack Overflow are closely related to the question-asker personally (e.g., asking the reasons for their own practices on Stack Overflow, such as posting or editing, being criticized and penalized), while other questions are based on their observations on the platform or the community and are only indirectly related to themselves. We use the notion of personal relatedness to capture this difference. Referring to CLT, questions that are highly personally related to the question-askers themselves can be mentally represented to be less psychologically distant than the questions that have a lower level of personal relatedness, and as a result, people may process these questions in a more intimate level.

Psychological distance is known to be closely related to emotional intensity, which could be reflected in the sentiment of language use when question-askers construct their posts. Research in emotion regulation shows that emotion intensity decreases as psychological distance increases, and that distancing from events can attenuate one's negative feelings [7, 10]. Since posting on Meta Stack Overflow is usually triggered by negative experiences or events on Stack Overflow, we pose H2a:

H2a. *Question posts that are personally related to the question-askers themselves have more negative sentiment compared to the questions that are not directly related to the question-askers.*

From the perspective of question-readers (who may also become question-commenters or question-answerer), questions that are

only related to the askers themselves may be perceived to be self-serving or not beneficial to a broader range of audiences. On the contrary, questions that are not directly related to a specific asker (for example, questions asked based on their observation in the community) are likely to be perceived as an act of helping someone else or contributing to the community. Posting questions that are less personally related, deemed as prosocial behavior, may draw more positive responses from the community. Thus, we pose H2b:

H2b. *Question posts that are personally related to the question-askers themselves receive more negative responses compared to the questions that are not directly related to the question-askers.*

2.2.2 Question specificity. Another difference in the framing of questions can be conceptualized as question specificity. One typical class of questions asking about community norms on Meta Stack Overflow targets a specific case or post on Stack Overflow, while the other type of questions are more abstract and do not refer to a specific case or post as the sole target of inquiry. Focusing on a specific case (which is usually about a negative experience on Stack Overflow) when framing a question causes the question-asker to think concretely about the event from a psychologically near perspective and "relive" the negative experience, which may lead to high emotional arousal [9]. Hence, we pose H3a:

H3a. *Concrete questions that target specific cases have more negative sentiment compared to the questions that are more abstract.*

Question posts that target a specific case on Stack Overflow may benefit the community very limitedly. In contrast, posting abstract questions that reflect on more general cases and cover a larger scope may be viewed as prosocial behavior because discussions around such questions are not only worthwhile for the question-askers themselves but also valuable to general audiences in the community. Therefore, we pose H3b:

H3b. *Concrete questions that target specific cases draw more negative responses compared to the questions that are more abstract.*

3 DATA AND METHOD

We conducted quantitative analyses based on the Meta Stack Overflow posts to test our hypotheses. In this section, we describe how we prepared the dataset.

3.1 Data Preprocessing

Mechanisms and community norms on Stack Overflow evolve constantly. In the interests of consistency, we only extracted posts content contributed to Meta Stack Overflow in 2021 (accessed through Stack Exchange Data Dump [7]) as the starting point since it was the most recent whole-year dataset at the time when this study was conducted. We filtered the question posts based on tags. Meta Stack Overflow requires each question to have at least one of the four

---

[3]https://meta.stackoverflow.com/questions/409150/does-images-of-code-denote-a-low-quality-question

[4]https://meta.stackoverflow.com/questions/413012/have-you-noticed-a-shocking-decline-in-question-quality-the-last-year-or-two

[5]https://meta.stackoverflow.com/questions/413266/why-did-my-question-get-closed-i-believe-its-high-quality-but-it-was-closed-do

[6]https://meta.stackoverflow.com/questions/413429/should-we-lock-questions-responses-from-downvotes-after-a-period-of-time

[7]https://archive.org/details/stackexchange. Source of the content comes from the Stack Exchange Network.

Personal relatedness	Personal	*I'm confused, I flagged [this question] as Low Quality but my flag has been declined by a moderator? [...]. If this doesn't denote a low quality question what does? According to [this meta question] it does, so why was my flag declined?* [3] *(user692942)*
	Non-personal	*I'm not sure if I am just growing exhausted, but it seems to me like the quality of questions has dramatically decreased in the last year or so. I assume this is because [...]. What do you think? What can we do about it?* [4] *(Libra)*
Question specificity	Specific	*I asked a question I thought was very good quality, and it received two downvotes and two close votes without any explanation. [...]. If someone could at least explain what is wrong with, for example, my question, that would be nice so I could correct it. [...].* [5] *(davidsbro)*
	Non-specific	*Is it possible, or even wise, to stop accepting downvotes on questions and/or responses after a set period of time? [...]. I agree downvoting is necessary and should inspire the author to take strides in improving their post. [...]. However, I see little value gained in downvoting a post months after it has been published.* [6] *(Paul Stoner)*

Table 1: Examples of question posts categorized based on personal relatedness and question specificity

required tags: `discussion`, `support`, `bug`, and `feature-request`. As our focus is on conversations about community norms which generally start with question posts that are open-ended in nature and do not have a concrete answer, we only kept question posts that have the `discussion` tag and are not contributed by anonymous users, leading to 1,215 question posts in total. We randomly picked 243 question posts among them (20% of qualified question posts) so that the size is manageable for manual coding required in a later stage.

Two researchers read all of them and manually excluded the question posts that are neither relevant to community norms nor likely to trigger discussions around norms in answer posts and comments. These question posts are then merged with their corresponding answer posts, comments, and user information. The body of the posts is preprocessed for better text readability (e.g., removing HTML tags). The final dataset includes 123 threads (consisting of 123 questions, 184 answers, and 1,737 comments).

3.2 Categories of Personal Relatedness and Question Specificity

Two researchers individually categorized all question posts based on their personal relatedness and question specificity, then inspected all initial labels together and discussed any differences in their labeling. Categorization was able to be finalized without any disagreement. Each question post has two labels: *Personal* or *Non-personal* (indicating personal relatedness), and *Specific* or *Non-specific* (indicating question specificity). We show example posts with the four labels in Table 1. Following the attribution requirement of the Stack Exchange Data Dump, we include the author names of the posts.

We performed sentiment analysis for each question post, answer post, and comment using VADER, a lexicon and rule-based model sensitive to sentiment expressions in social media text [5]. We used the output compound score, ranging from −1 (most extreme negative) to +1 (most extreme positive), as our measure of the sentiment of each individual post.

4 DATA ANALYSIS AND RESULTS

4.1 Question-askers' Experience Levels and the Questions' Well-receivedness

We used question-askers' reputation score as the measure of their experience levels, and we consider questions that have higher view count, answer count, comment count, and score (i.e., votes received) to be more well-received. Fig. 1 shows the distribution of the 123 question-askers' reputations. Pearson's correlation test did not indicate reputation to be correlated with view count ($r = 0.088$), answer count ($r = -0.036$), comment count ($r = 0.004$), and score ($r = -0.138$). H1 is not supported.

While many previous studies showed that novices are disadvantaged on Stack Overflow and questions asked by more experienced users are likely to be more well-received, we did not see a similar pattern on Meta Stack Overflow. This might be caused by the reputation-gated permission on Meta Stack Overflow. While anyone can ask a question on Stack Overflow, only users who have at least a reputation score of 5 are allowed to create a new question post (with some exceptions) [8]. The non-existence of users who have completely no experience in our Meta Stack Overflow dataset may be one reason that we did not observe a similar effect between Meta Stack Overflow and Stack Overflow. From another perspective, knowledge gaps between experienced users and novice users on Stack Overflow can fall into at least two categories: the gap in technical knowledge (e.g., the unfamiliarity with a software package) and the gap in the platform- or community-related knowledge (e.g., how to create a good question post). While both of these gaps could play essential roles in Q&A on Stack Overflow, the gap in technical knowledge is not very relevant in conversations around community norms on Meta Stack Overflow. In this sense, when it comes to norms-related discussions, we can speculate that the gaps between experienced users and novices are "mitigated", which could be another possible reason that users' experience levels do not seem to play important roles in the questions' well-receivedness on Meta Stack Overflow.
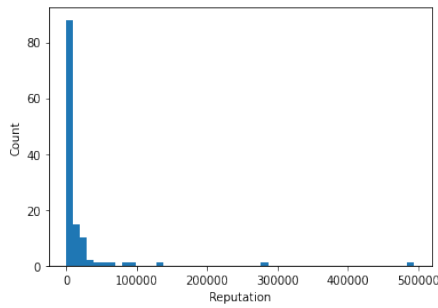
---

[8]https://meta.stackoverflow.com/questions/ask

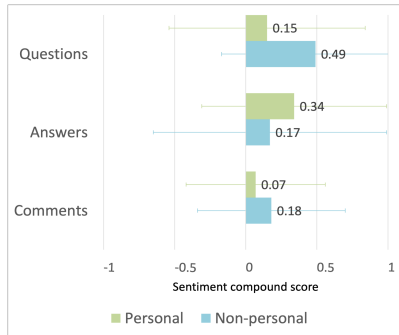**Figure 1: Distribution of the 123 question-askers' reputations.**

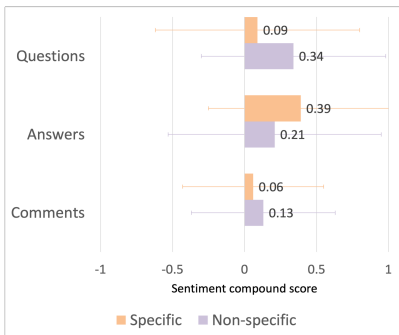**Figure 2: The effects of personal relatedness.**

Figure 3: The effects of question specificity.

4.2 Personal Relatedness and Conversational Sentiment

T-tests were performed to analyze the effect of question posts' personal relatedness on the conversation sentiment. While not statistically significant, we found that the sentiment of personally related questions is more negative than questions that are not personally related ($t(121) = -1.91, p = 0.07$; Personal: $M = 0.15, SD = 0.69$; Non-personal: $M = 0.49, SD = 0.66$) (shown in Fig. 2). We found no significant difference in sentiment between answers to personally related questions and non-personally related questions. Compared to the sentiment of comments of personally related questions, the sentiment of comments to non-personally related questions is significantly more positive ($t(1735) = -3.16, p < 0.01$; Personal:

$M = 0.07, SD = 0.49$; Non-personal: $M = 0.18, SD = 0.52$). In summary, H2a is supported and H2b is partly supported. One possible explanation that there is a significant sentiment difference in comments, but not in answers, is that answers tend to be emotionally neutral while comments usually include more subjective expressions.

4.3 Question Specificity and Conversational Sentiment

We ran t-tests to compare the sentiment in conversations initiated by question posts with different levels of question specificity. We found that sentiment in questions that have a specific target is more negative than sentiment in questions that are more abstract ($t(121) = -2.09, p < 0.05$; Specific: $M = 0.09, SD = 0.71$; Non-specific: $M = 0.34, SD = 0.64$) (see in Fig. 3). There is no significant difference in sentiment between answers to questions with a specific target and questions without a specific target. Statistics showed evidence that sentiment in comments to relatively abstract questions (in other words, questions that are relevant to a larger scope) is significantly more positive than comments to questions with a specific target ($t(1735) = -2.74, p < 0.01$; Specific: $M = 0.06, SD = 0.49$; Non-specific: $M = 0.13, SD = 0.50$). These results support H3a and partly support H3b. Similar to the effect of personal relatedness, the different effects of question specificity on answer posts and comments may also be caused by the relevant objectivity of answer posts compared to the comments.

5 SUMMARY

In large-scale online communities that employ complex mechanisms, designing effective communication channels for people to discuss community norms is essential for keeping the communities inclusive. In this paper, we focused on discussions around community norms on Meta Stack Overflow and investigated the influences of question askers' experience levels, questions' personal relatedness, and question specificity on the community's reactions and conversational patterns. We found that personal relatedness and question specificity of question posts on Meta Stack Overflow have main effects not only on the sentiment in the body of the question posts themselves but also on the sentiment in the comments they received. Our findings align with social psychology research in construal level theory and emotion regulation, which reveal that questions that are less personally related and questions that are asked on a broader scope (rather than narrow down the target to a specific case) have more positive sentiments. In addition, such question posts, which are likely to be deemed prosocial by the community, also receive comments with more positive sentiments. Our study also implies that online community maintainers can establish a healthier conversational environment for norms-related discussions by designing mechanisms that encourage non-personal and non-specific posts.

While a quantitative approach can be a starting point for studying conversations around online community norms, we plan to conduct a more in-depth content analysis on the textual data to further unfold the discussions and derive more insights into how conversational dynamics are influenced by different strategies used

in framing the posts. Future work can also investigate how people navigate discussions around norms in non-technical and non-knowledge-based online communities.

REFERENCES

[1] Muhammad Asaduzzaman, Ahmed Shah Mashiyat, Chanchal K Roy, and Kevin A Schneider. 2013. Answering questions about unanswered questions of stack overflow. In *2013 10th Working Conference on Mining Software Repositories (MSR)*. IEEE, 97–100.

[2] Amy Bruckman. 1998. Community support for constructionist learning. *Computer Supported Cooperative Work (CSCW)* 7 (1998), 47–86.

[3] Denae Ford, Kristina Lustig, Jeremy Banks, and Chris Parnin. 2018. " We Don't Do That Here" How Collaborative Editing with Mentors Improves Engagement in Social Q&A Communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[4] Denae Ford, Justin Smith, Philip J Guo, and Chris Parnin. 2016. Paradise unplugged: Identifying barriers for female participation on stack overflow. In *Proceedings of the 2016 24th ACM SIGSOFT International symposium on foundations of software engineering*. 846–857.

[5] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

[6] Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.

[7] Ethan Kross, Ozlem Ayduk, and Walter Mischel. 2005. When asking "why" does not hurt distinguishing rumination from reflective processing of negative emotions. *Psychological science* 16, 9 (2005), 709–715.

[8] Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design lessons from the fastest q&a site in the west. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2857–2866.

[9] Tal Moran and Tal Eyal. 2022. Emotion regulation by psychological distance and level of abstraction: Two meta-analyses. *Personality and Social Psychology Review* 26, 2 (2022), 112–159.

[10] Andreas Mühlberger, Roland Neumann, Matthias J Wieser, and Paul Pauli. 2008. The impact of changes in spatial distance on emotional responses. *Emotion* 8, 2 (2008), 192.

[11] Rogier Slag, Mike de Waard, and Alberto Bacchelli. 2015. One-day flies on stackoverflow-why the vast majority of stackoverflow users only posts once. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*. IEEE, 458–461.

[12] 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*. 1379–1392.

[13] Yaacov Trope and Nira Liberman. 2010. Construal-level theory of psychological distance. *Psychological review* 117, 2 (2010), 440.