Corrective or Critical? Commenting on Bad Questions in Q&A

Andrew W. Vargo[1], Shigeo Matsubara[1]
[1]Kyoto University

Abstract
What kind of comments do Q&A community members prefer on bad questions? We studied this question on Stack Overflow, a large-scale community-run question-answer site with strong reputation and privilege systems and clear guidelines on commenting. Peer-production systems often employ feedback dialogue to engage with producers of low quality content. However, dialogue is only beneficial if it is constructive, as previous work has shown the adverse effects of negative feedback on quality and production. Previous studies indicate that feedback is likely critical, but the extent, orientation, and actors within this assumption are unknown. In this paper, we contribute a basic taxonomy of commenting and perform analysis on user types and community preferences. Results indicate that the most popular and frequent comments include criticism, and that different user types leave similar feedback. A better understanding of community feedback norms can inform the design of effective rules and incentives for peer-production systems.
Keywords: Q&A; online community; community feedback systems; qualitative methods
doi: 10.9776/16199

Acknowledgements: The authors would like to thank Naomi Yamashita and Ari Hautasaari at NTT Communication Science Laboratories for their encouragement and invaluable feedback on this manuscript. This research was partially supported by a Grant-in-Aid for Scientific Research (S) (24220002, 2012-2016) from Japan Society for the Promotion of Science (JSPS).
Contact: awv@ai.soc.i.kyoto-u.ac.jp

1 Introduction

The introduction of domain specific corpus-building question-answering (Q&A) has enabled professionals, experts, and students of many different subjects to solve problems and tasks of varying difficulties with great speed and effectiveness (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2012; Mamykina, Manoim, Mittal, Hripcsak, & Hartmann, 2011). These communities are capable of not only producing timely answers; they are also adept at creating archives with lasting impact (Anderson et al., 2012) and are able to collaborate on difficult technical problems (Tausczik & Pennebaker, 2012). A problem that exists for successful sites is that the number of poor questions will start to hinder the efficacy of the archive, and an effective way of handling this problem is to let the community moderate the content itself (Correa & Sureka, 2014). However, as effective as the process is at removing unwanted questions, one concern facing such systems is the quality of feedback dialogue that is given to the contributors of the question. In this research, we analyze how a community comments on unwanted material, and how the community judges different types of feedback through a voting mechanism.

Dialogue based feedback on contributions has been often studied in different peer-production communities. Empirical studies have shown that the type of feedback that users receive in peer-production systems can affect their following contributions. Negative feedback can decrease a user's future contributions (Zhu, Kraut, & Kittur, 2012), while positive feedback can encourage new users to contribute more (Choi, Alexander, Kraut, & Levine, 2010). In social media systems, negative feedback can create a loop in which the recipients of negative feedback produce more poorly rated material and become negative themselves (Cheng, Danescu-Niculescu-Mizil, & Leskovec, 2014).

A limitation of these studies is the focus on correlative value and impact on the receiver of the comment (Zhu, Zhang, He, Kraut, & Kittur, 2013), rather than a focus on the creators of the feedback. It is unclear whether contributors of critical feedback are trolls, low reputation users, or the entire community. In addition, studies that draw correlation based off data-mining (Ahn, Butler, Weng, & Webster, 2013; Zhu, Kraut, Wang, & Kittur, 2011) are limited by a lack of context and nuance. It is difficult to rely on machine mining of negative comments, as kappa agreement with human evaluators is low to moderate (Zhu et al., 2011). This signifies that research can be strengthened by qualitative methods.

In this research, we study how the community interacts with bad questions in Stack Overflow (SO), a Q&A system dedicated to computer programming problems through commenting. SO is an appropriate venue for this research because there are thousands of questions each day, and there are numerous questions that fail to meet the question guidelines, and as a result, there is a necessity for community moderation. Likewise, comments are both regulated by a series of rules that encourage constructive feedback, and comments receive votes, which reveals the preferences of the community.

SO also is governed by a reputation and privilege system that is based on the idea that reputation points are synonymous with trust. SO users with enough reputation points are given moderating powers. Previous work has indicated that volunteer administrators on Wikipedia will use more positive emotions in discussion pages (Laniado, Kaltenbrunner, Castillo, & Morell, 2012), so there is potential for experienced community members to guide inexperienced or poor questioners to fix their errors and improve future contributions through constructive dialogue and feedback. In the same way, unconstructive criticism could potentially come from lower reputation users.

In this paper, we extend the understanding of feedback and its relationship with a community by performing a qualitative review. We developed a basic taxonomy through a Grounded Theory process (Glaser & Strauss, 2009), with special focus on the context of the comment with the question. The process allows the research to take into account the question and context of the feedback. We then investigated whether the community preferred certain types of comments by using votes given to comments, and whether different types of users, especially users with high reputation, contribute differently from each other. Lessons from this study are useful for better understanding community orientation towards feedback in Q&A and peer-production communities, as well as informing rules and incentives for these systems.

2 Related Work

There are two areas of related work that we address. First, we look at previous work related to commenting and feedback systems. Then we outline work that has been conducted on the motivations for contributors in Q&A and Stack Overflow (SO).

2.1 Feedback and Learning in Peer Production Communities

Wikipedia is a standard-bearer for large-scale peer-production systems, and relies on feedback in order to create high quality articles (Laniado et al., 2012). A large number of peripheral editors play a significant role in providing leadership and feedback (Zhu et al., 2011). In order to achieve a continuous critical mass of content, educating newcomers to the community is essential (Schneider, Passant, & Decker, 2012). Two studies (Halfaker, Kittur, & Riedl, 2011; Zhu et al., 2012) found that negative feedback, through reversions and commenting, harms new participants and fails to improve future outcomes, as opposed to positive feedback. Zhu (Zhu et al., 2013) did find, however, that in lab controlled settings, mild negative feedback did have a corrective influence on participants.

Negative feedback can also mean more poor contributions in social computing situations. Cheng (Cheng et al., 2014) reviewed data from online news communities, and found that negative feedback actually propagated more negatively received contributions, while positive feedback did not evoke a similar response from users. On the other hand, Wohn (Wohn, 2015) found in the social media site Everything2 that negative feedback could have a positive effect on established users who had developed a habit and perhaps should be used for experienced users only. These studies on Wikipedia and the other online communities do indicate that negative feedback does not necessarily have a deterministic outcome. There are some situations in which negative feedback could be useful for a community's aims. In order to understand outcomes beyond correlation, it is important to understand how the feedback is actually formulated and propagated by the community.

There has also been some research regarding feedback on SO and its network of Q&A sites, Stack Exchange. Tausczik (Tausczik, Kittur, & Kraut, 2014) found that commenting could be used to help solve complex problems on the Stack Exchange site Math Overflow, which indicates that commenting has a tangible use in this type of system. Yang (Yang, Hauff, Bozzon, & Houben, 2014) explored the theoretical possibilities that commenting could be used as part of editing to moderate poorly formulated questions. However, Ahn (Ahn et al., 2013) found that the volume of comments had no positive effect on the future participation of a questioner. The authors concluded that content and not volume may play the more pivotal role in influencing future behavior. Li (Li, Zhu, Lu, Ding, & Gu, 2015) looked at constructive commenting as a function of collaborative editing, and found that commenting did have a small negative effect on the future participation of the questioner or answerer. Since the role of feedback and its effect on users is relatively clear, in this paper, we do not focus on the correlative nature of feedback on future outcomes, but rather, on the contributions and interactions themselves as part of its role in the community.

2.2 Motivations to Contribute in a Q&A Community

Why do users and experts contribute and expend effort in social Q&A? The first reason may simply be an attachment to the site. In sites like SO, a comprehensive outreach and discussion with a domain community can create strong ties to the community (Mamykina et al., 2011; Tausczik & Pennebaker, 2012). The role of social approval is clearly important. Users tend to give answers on content in which they feel that their effort can be recognized by the questioner and the broader community, and will avoid questions that are too narrow or broad (Anderson et al., 2012; Pal & Konstan, 2010; Treude, Barzilay, & Storey, 2011). In general, users will contribute more when they receive social approval through feedback from the community (Cheshire, 2007; Tausczik & Pennebaker, 2012). Social approval is usually considered from the Q&A activity itself. We look at social approval's impact in a non-reputation earning context.

Gamification can also motivate contributors. Reputation points are often associated with signifying experts and encouraging quick answers. The drive to earn reputation is strong enough that the reputation points incentivize users whether they recognize it or not (Tausczik & Pennebaker, 2012). This also correlates strongly with the behavior of experts, who target the most reputation valuable questions (Anderson et al., 2012). In addition to earning reputation points, badges have been shown to significantly encourage desired Q&A behavior (Cavusoglu, Li, & Huang, 2015) and support activities like editing (Grant & Betts, 2013). In this paper, we do consider that gamification may play an auxiliary role in commenting behavior.

3 Stack Overflow Overview

3.1 System Overview

By the end of 2013, Stack Overflow (SO) had over 2.7 million registered users who had asked 6.5 million questions and have provided 11 million answers. Users on SO are differentiated via an interactive reputation system. Community members can earn and lose points from votes on their questions and answers. They can, in turn, vote on other users' contributions. Votes from answering questions are 10 points and are the most common source of points (Anderson et al., 2012).

While the community is large, it does have a steep reputation pyramid, in which a handful of expert users own most of the reputation (Anderson et al., 2012). This is not uncommon for Q&A (Pal, Chang, & Konstan, 2012). This is important because reputation points earn users privileges such as voting, closing, and commenting. For instance, only 1% of users can participate in the closing of questions, as shown in Table 1. Thus, the system is strongly hierarchical.

Privilege	Points Needed to Obtain Privilege	Users with Privilege
Vote Up	15 Points	24%
Comment	50 Points	12%
Vote Down	125 Points	8%
Vote to Close Questions	3000 Points	1%

Table 1. Related Privileges on Stack Overflow

3.2 Asking Questions

SO is very particular with regard to the types of questions it allows on the site. This is in order to avoid the problem of having too many questions and not enough answers (Guo, Xu, Bao, & Yu, 2008). SO was explicitly designed with the idea that questions are easier to ask than answer; and that too many questions, or poor questions, would lead to an unsustainable eco-system.[1] Part of this engineering is implemented by setting rules, expectations, and limitations on what can be asked.

Anyone can sign up for an account on SO and ask a question to the community in a matter of minutes. Of course, new questioners may be unprepared for asking questions that are appropriate for the forum. When a new user signs up to ask a question, a pre-form guide to asking a question greets them.[2] This guide highlights certain steps that a user should take before asking the question. Notably, SO asks that questioners perform and show prior research, stick to the topics of the site, and make the question relevant to others. New users should check the box at the bottom of the page to confirm that they will

[1] http://ethnographymatters.net/blog/2012/02/24/online-reputation-its-contextual/
[2] http://stackoverflow.com/questions/how-to-ask [August 2013 archive accessed September 1, 2015]

follow these rules. Included in this page is a number of links to guides on how to ask and how not to ask questions.

3.3 Process of Commenting on Bad Questions

SO has a clear guideline for the commenting privilege, most importantly that comments should ask for clarification or give constructive feedback that helps the author improve the post. On the other hand, criticism which is not constructive or answering a question is not allowed.[3]

Figure 1 shows an example of the commenting process on SO. At 10:50, five minutes after the question was posted, User A comments on the simplicity of the question and motivation of the questioner by writing, “people are so lazy to make some googlefu”. Later, at 11:02, User C comments on the question, but in a completely new direction, instead explaining to the questioner why the question was poorly received. The questioner responds to User C with thanks for the guidance.

The votes on the comments do not count towards reputation, but they do act as a form of social approval since the community clearly states its preferences (Cheshire, 2007). While there has been a discussion to make these comments count towards reputation, this has been rejected by the community.[4] However, users can earn a participation badge for comments receiving more than 5 votes, and the voting structure does indicate the community’s preferences. In this case, we can see that User A’s comment was the most popular with 11 votes, while User C’s comment was second with 5 votes.

The example in Figure 1 shows an interesting dynamic. The initial comment contributes a harsh criticism of the questioners’ abilities and motivations. The third comment contributes guidance. This example shows how the contribution of commenters can vary greatly from user to user, and informs this research: should we expect commenters to contribute constructive criticism and guidance, or is the ratio more skewed toward criticism without guidance?

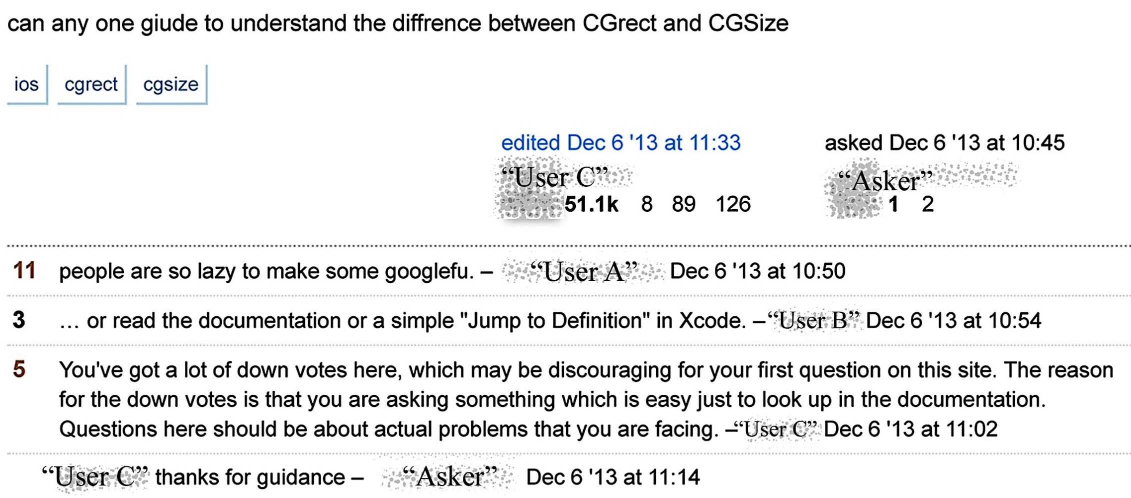

Figure 1. Question 20421944: An Example of a Bad Question and Comments (Anonymized).

The prevalence of unfriendliness was concerning enough for SO’s administration that the CEO and co-founder, Joel Spolsky, launched a campaign called the, “kicking off the summer of love” in 2012.[5] A key passage from the blog post stated:

> Newbies will show up, make a newbie mistake, like wearing shoes indoors or forgetting to close the toilet lid, and the old-timers will look at each other, roll their eyes, and snort, "Typical!" At this point, if it's a normal human community, it will start to feel a little bit unfriendly to outsiders. Insular. And the newbies will say, "well, gosh, that's not a very friendly place." Not just the newbies who got scolded. Also the 100 passers-by who *saw* the newbies get scolded. Who might have been great members of the community, and who did nothing wrong, but who are not really interested in joining a community that appears to be full of smug jerks. This is very dangerous. You have to be able to recruit new members to replace the old ones that drift away. The success of the community depends on it.

[3] http://stackoverflow.com/help/privileges/comment [August 2013 archive accessed September 1, 2015]
[4] http://meta.stackexchange.com/questions/17364/how-does-comment-voting-and-flagging-work
[5] https://blog.stackexchange.com/2012/07/kicking-off-the-summer-of-love/

From this point, the community leaders made concerted efforts to educate community members on what constitutes acceptable feedback.[6][7] Specifically, tone and content were singled out, asking users to be civil and informative.

4 Data Set

Since the goal of this study is to look closely at interaction that occurs on *clearly* bad questions on Stack Overflow (SO), we created a number of rules to collect a sample that would allow for a close investigation: 1) Questions must have more than five down votes: The rationale behind this rule is to make sure that a larger section of the community agrees with the closers that the question is not suitable for SO, 2) Questions must be asked by a unique user in order to be included in the data set. There are no repeated questioners, and 3) Questions must have at least one comment that is made by someone other than the author.

From these rules, we were able to identify approximately 20,000 eligible questions. We took a random sample of N=581 (5% margin of error, 98% confidence). Some questioners were completely new (N=241), while other had asked before (N=340). The mean of previous questions was 18.4 (SD=97.9, Mdn=1). The mean for the number of comments was 4.5 (SD=3.5, Mdn=4).

The commenters came from various backgrounds as well. We used the classification scheme developed by SO to classify the users. SO defines users on the following metrics: User – 0 to 999 reputation points, Established User - 1,000 to 19,999 reputation points, and Trusted User – 20,000+ reputation points.[8] There were seven commenters whose reputation could not be identified and are not included in Table 3. As mentioned previously in Table 1, reputation points earn users privileges. Trusted Users have all privileges, including significant editing and voting powers. As we mentioned earlier, reputation in SO is supposed to be a rough measure of trust a user has in the community, and the system gives privileges based on this trust. From a system point of view, Trusted Users would be expected to closely follow the commenting rules.

Reputation	Number of Commenters	% of Comments	Unique Commenters
User	108	18.6%	107
Established User	284	48.9%	254
Trusted User	182	31.3%	121

Table 3. Number of Commenters by Reputation Class

5 The Formulation of Bad Questions

An important part of understanding the interaction on these questions is having a clear view of what these questions are trying to accomplish. We used a Grounded Theory approach (Glaser & Strauss, 2009) to distill comprehensible categories. We started with a simple question in order to orient the study, “What is the questioner trying to achieve by asking the question?” Instead of trying to understand why the question was good or answerable, we focused on the end goal of the questioner. The open-coding session led to the identification of a small number of strong themes that quickly coalesced into four distinct categories, which align with previous research categorizations (Treude et al., 2011):

- **Process:** These questions ask for help in completing a process. These questions are often when a problem has been setup with code, and the questioner does not know the next step in order to achieve an objective. (152 questions)
- **Process without code:** These questions are essentially the same as Process questions, but lack any code or fail to show any attempted work. (163 questions)
- **Debugging**: The questioner has a problem or error in code. (186 questions)
- **General Concept**: These questions deal with choosing a tool or theoretical questions such as, “what language is best to learn for a future career?” (80 questions)

[6] https://blog.stackexchange.com/2012/08/stack-exchange-is-not-a-forum-the-role-of-niceness-on-a-qa-site/
[7] http://meta.stackexchange.com/questions/138173/etiquette-for-posting-civil-and-informative-comments/
[8] http://stackoverflow.com/help/privileges [August 2013 archive accessed September 1, 2015]

6 How does the Community Comment on Bad Questions?

6.1 Methods

A grounded theory approach (Glaser & Strauss, 2009) was used to inductively categorize the type of contributions that questioners receive from commenters. The entire question set was studied in detail by the authors in order to understand the content given by the commenters. In this way, the coding was conducted from the viewpoint of a questioner. There were two types of coding: open coding, where the categories and sub-categories were discussed and mediated, and individual coding, where the same coders independently assigned values to denote strength of a category within a comment. During the open-coding session, numerous labels and sub-categories were identified. These sub-categories were then used to guide the formation of more abstract categories. In total, we were able to distill three distinct categories of comments.

- **Corrective:** These comments explained procedural elements of why the question is bad or gave explicit help on how to improve the question. For instance: providing reference to Stack Overflow (SO) questioning procedures; giving direct guidance to asking a better question; explaining the reaction of the community.
- **Critical:** These comments point out flaws in the question and are unconstructive. For instance: giving accusations of a lack of effort; of cheating on homework; of a lack of ability; of trying to get work done for free; or mocking user's ability.
- **Answer:** These comments addressed the question. For instance: giving a direct link to answer; giving correction of error in a question; giving a full answer.

Question Type	Comment	Type of comment
Process without code	You've shown no effort, are basically asking for someone to do the work for you.. and you've tagged this with tags that don't relate to SQL..	Critical: Accusation of lack of effort; Lack of skill Corrective: Gives correction on content
Process	"I have a class which represents database record with over 100 string fields" - OMG you have a much greater problem than string initialization.	Critical: Lack of Skill
Debugging	Because it prints, then increments. ++index will increment, then print. Postfix versus prefix.	Answer: Comment explains the error in a short comment
General Concept	Which century are you living in?	Critical: Lack of skill;
Process	Welcome to Stack Overflow. Please read the FAQ on how to ask questions here - you'll need to describe what you've tried so far and what didn't work, preferably with code samples. At that point we're more than happy to help.	Corrective: Gives directions on how to ask

Table 4. Examples of Comments and Codes

The interactions present in the comments showed a wide range of variability. Corrective comments often focused on the specific process that needs to be followed. For instance, in Table 4, the last example comment refers to the FAQ that the questioner has already agreed to. In this way, we found that corrective commenting often gave a pedagogical guide to the questioner. However, correction also occurred in other ways, when errors in the formatting or general guidance were given along with criticism.

Criticism was often given in a way that either directly accused the user of a misdeed, or mocked their question. The first comment gives an example of a direct accusation of being lazy and violating the spirit of the site. Comments two and four, on the other hand, give less direct but strong criticism regarding the questioners' skill and understanding of the content. To a certain extent, these comments would seem completely in violation of the commenting rules as they fail to be constructive at all.

6.2 Coding Categorization of Comments

In order to better understand how an entire comment is read in the context of the question and to understand the distribution of comments, we employed a taxonomic coding scheme based on the open coding results to categorize the comments as received by the questioner. To initiate coding, the independent coders participated in open coding of the contributions. Disagreements on definitions were resolved through discussion. The coders rated a selection of the top scoring comments (N=581) from the data set. We chose this method because this study focuses not of the popularity of commenting on certain types of bad questions, but on the popularity of commenting types. Rating the top comment as opposed to all comments removes homophily that can be present in discussions and gives a clearer picture of what kind of comments are popular. When there was a tie (N=110) the earlier comment was selected for coding. Three coders, including the first author, participated in the coding. The first coder has a Masters in Information Science and the second is a Masters student in Media Studies. While all of the coders have familiarity with computer programming topics, none of them has been or are active users of SO.

Comment Type	Presence in Comments with a Score of ≥1	Comment Mean and Standard Deviation	Comments with an Isolated Score of ≥1	Inter-rater Agreement (ICC)
Corrective	47.1%	M=0.92 SD=0.98	13.23%	0.78
Critical	66.1%	M=1.57 SD=1.52	35.1%	0.72
Answer	28.1%	M=0.62 SD=0.96	16.0%	0.84

Table 5. Comments by Classification

The coders were given the comment with the original question. In addition, the coders were allowed to follow the off-site links to view resources that were mentioned in the comments. The reputation of the commenters or questioners was not available. The coders were asked to rate the comments on the following five-point Likert scale (from extensively (4) to not at all (0)) for each of the three categories. The coders considered each contribution on these evaluation points. A comment could have the elements of multiple categories. An important concern before the coding session was that of tone and implied content in the comments could make it difficult to obtain high agreement. However, they were not guided to ignore tone or implied content due to the design of the test.

We calculated the inter-rater agreement as a Cohen's kappa using Shrout and Fleiss ICC2 as shown in Table 5, and the results show that there was strong agreement. Comments tended to have strength in more than one type of category, explaining why total presence of comments exceeds 100%.

We also looked at isolated scores. Isolated scoring highlights a comment that may have more than one type of attribute, but is stronger in one particular area. This is done through subtracting aggregate scores of each category. If a comment received aggregate scores of 1 for corrective, 1 for critical, and 3 for answer, it would be classified as an answer comment with an isolated score of 1. Corrective comments were the least likely to be isolated, while critical comments were the most likely.

	No. of Comments	Score ≥ 1	No. of Comments	Isolated Score ≥ 1
Corrective	274	3.0073	77	2.8701
Critical	384	3.6354	204	4.1029
Answer	163	2.9202	93	2.9570
F-Value		6.2126**		7.4783**
Tukey A Vs B		4.1744**		4.4114**
Tukey A Vs C		0.4625		0.2698
Tukey B Vs C		4.0207*		4.3835**

Table 6. ANOVA: Comments by Type and Votes Received (*p<0.05, **p<0.01)

Votes on Comments: Next, we looked at the number of votes received on different types of comments. First we looked at all comments that received an aggregate score of ≥1 in a particular category. An ANOVA with post-hoc Tukey (as shown in Table 6) showed that critical comments received higher scores than both corrective and answer comments. The isolated critical comments are even more popular as they gain about 4.1 votes on average. This would indicate that the community has a strong preference for criticism. An important thing to note is that this analysis only looks at comments that received the most votes (or were tied for such status). It does not take into account comments that follow in the ratings.

Relationship with Commenter Reputation Class: The next thing we wanted to look at was the type of comments left by users from different reputation classes. An assumption could be made that Trusted and Established users would be more likely to follow the rules since they have been through the privilege process and may be more sensitive to the actions that are deemed socially desirable by the SO administration. As Table 7 shows, there was no significant difference between any of the classes and the type of comments left. Critical comments were almost identical between all classes.

A limitation of this test, however, is that we only look at the top rated comment. There is the possibility that Trusted Users give constructive comments that are not highly rated by the community members. However, it is important to recognize that while Trusted Users account for 31% of the comments, they are less than 0.1% of the community at large. In addition, Established Users are part of the top 1% of users, but 49% of all comments. The results thus can be considered to be indicative of a larger pattern of behavior.

	Corrective	Critical	Answer
User	0.8056	1.5123	0.6235
Established User	0.9437	1.5716	0.5892
Trusted User	0.9652	1.5513	0.6923
F-Value	1.0105	0.1040	0.6345

Table 7. ANOVA: Comments by Type and Commenter Reputation Class (*p<0.05, **p<0.01)

6.3 Summary and Discussion

We were able to identify three common types of comments that were left as popular feedback on the bad questions. The most common was critical, followed by corrective, and comments that helped to answer the question. We might expect that the most common type of feedback would receive the highest amount of voting, and this is true. Critical comments receive more votes than other types of comments. In addition, comments that consisted of only criticism received even more votes than other isolated types of comments. Finally, we found that comments came from all types of users, including high reputation users. The results not only confirm the suspicion from previous work (Ahn et al., 2013) that there is a critical trend in the comments, but that this trend is popular and affirmed within the community.

The results indicate, however, that social approval is given to criticism, some of which would be considered “unconstructive” by the SO administration. Since we can expect users will develop contribution patterns when they are socially affirmed by the community (Cheshire, 2007), this offers an explanation for why users of all reputation classes act similarly. Rather than criticism being something that happens by individual actors, it is part of a larger social construct. This is something that is cause for concern from a system design point of view as strong negativity has a likely adverse effect on producers and potential producers (Zhu et al., 2013), which might have a critical externality effect on outside users, and is specifically banned by the community administration. At the same time, there is a large amount of corrective commenting that is present, and it does receive some social approval. This indicates that there are users that support the socially desirable actions as defined by the administration.

7 Conclusion

In this paper, we examined the characteristics of bad questions and analyzed how the community of Stack Overflow (SO) comments on them. Questions usually seek help in completing some sort of task, such as creating a process or debugging code. We were able to determine that comments had large amounts of corrective guidance and criticism. Comments also sometimes aimed at giving an answer to question. We found that critical comments were the most popular and received the highest vote counts, and that commenters did not change their commenting behavior based on their reputation.

One observation that comes out of this research is that the community users have norms that differ from the administration. Q&A has expanded from generalist systems with little concern for archival efficacy to domain specific sites that seek to be an authoritative and useful archive of knowledge (Anderson et al., 2012). A challenge to these archives is the sheer number of bad questions that appear. In such situations, it is likely a natural outcome that the community both writes and approves of feedback that is critical and unconstructive. It should be noted, however, that not all of the comments contain negativity, and that there is a considerable amount of corrective feedback mixed into the comments. There is potential to draw corrective feedback to the forefront of the contributions.

The driving force behind the social approval of criticism may be that it essentially accomplishes a desired goal: to firmly chastise users who post bad questions. The interactions we reviewed show a strong emphasis on effort and skill. The commenters do emphasize that learning through reading and study is an essential part of learning to be a computer programmer. In addition, those who show no effort in their questions are despised. Commenters may believe that bad questioners have no potential value to the community, so critical comments are valid from the viewpoint of the community, regardless if it is actually socially beneficial from the viewpoint of the administrators.

One possibility for ameliorating the power of social approval is to create badges that support socially beneficial behavior. While we cannot conclude that the current commenting badge is directly responsible for unconstructive criticism, we can conclude that current voting system allows for the earning of these badges. This defeats the purpose of comments in situations like bad questions. The dialogue is supposed to support the questioner, who will then benefit the community by producing better content. A way to encourage this is to create badges that are directly linked to support situations. For instance, allowing questioners to select the most helpful comment and rewarding badges for such an effort could offer a tangible incentive to leave constructive feedback. Another recommendation is to focus on Trusted Users. Since previous work has identified that users are especially effected by the actions of core members (Zhu et al., 2013), focusing on these users' comments could have a beneficial network effect.

7.1 Limitations and Future Work

In this paper we looked at how a Q&A community gives public feedback to poor quality questions. The observations were gathered in a single system that has a technical focus and the study analyzed a subset of questions. Future work should look at not only other domains, but also systems of smaller sizes. Another possibility is looking at how commenting changes as the quality of respective questions change, or how commenting changes depending on the question contributor's status in the community. Such a study could be useful for indicating when the community is moved to be more critical and less corrective.

A limitation of this work lies in the lack of interviews. It would be informative to get the perspective of users who have contributed bad questions as well as gain an understanding of the explicit motives of users who comment on these questions. It would also be useful to understand if there is a pattern of commenting from contributors that evolves over time as they are exposed to more content.

8 References

Ahn, J., Butler, B. S., Weng, C., & Webster, S. (2013). Learning to be a better q'er in social Q&A sites: Social norms and information artifacts. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1–10. http://doi.org/10.1002/meet.14505001032

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2012). Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 850–858). New York, NY, USA: ACM. http://doi.org/10.1145/2339530.2339665

Cavusoglu, H., Li, Z., & Huang, K.-W. (2015). Can Gamification Motivate Voluntary Contributions?: The Case of StackOverflow Q&A Community. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (pp. 171–174). New York, NY, USA: ACM. http://doi.org/10.1145/2685553.2698999

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2014). How Community Feedback Shapes User Behavior. In *Proceedings of ICWSM*. Retrieved from http://arxiv.org/abs/1405.1429

Cheshire, C. (2007). Selective Incentives and Generalized Information Exchange. *Social Psychology Quarterly*, *70*(1), 82–100. http://doi.org/10.1177/019027250707000109

Choi, B., Alexander, K., Kraut, R. E., & Levine, J. M. (2010). Socialization Tactics in Wikipedia and Their Effects. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 107–116). New York, NY, USA: ACM. http://doi.org/10.1145/1718918.1718940

Correa, D., & Sureka, A. (2014). Chaff from the Wheat: Characterization and Modeling of Deleted Questions on Stack Overflow. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 631–642). New York, NY, USA: ACM. http://doi.org/10.1145/2566486.2568036

Glaser, B. G., & Strauss, A. L. (2009). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Transaction Publishers.

Grant, S., & Betts, B. (2013). Encouraging user behaviour with achievements: An empirical study. In *2013 10th IEEE Working Conference on Mining Software Repositories (MSR)* (pp. 65–68). http://doi.org/10.1109/MSR.2013.6624007

Guo, J., Xu, S., Bao, S., & Yu, Y. (2008). Tapping on the Potential of Q&a Community by Recommending Answer Providers. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (pp. 921–930). New York, NY, USA: ACM. http://doi.org/10.1145/1458082.1458204

Halfaker, A., Kittur, A., & Riedl, J. (2011). Don'T Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 163–172). New York, NY, USA: ACM. http://doi.org/10.1145/2038558.2038585

Laniado, D., Kaltenbrunner, A., Castillo, C., & Morell, M. F. (2012). Emotions and Dialogue in a Peer-production Community: The Case of Wikipedia. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (pp. 9:1–9:10). New York, NY, USA: ACM. http://doi.org/10.1145/2462932.2462944

Li, G., Zhu, H., Lu, T., Ding, X., & Gu, N. (2015). Is It Good to Be Like Wikipedia?: Exploring the Trade-offs of Introducing Collaborative Editing Model to Q&A Sites. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1080–1091). New York, NY, USA: ACM. http://doi.org/10.1145/2675133.2675155

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design Lessons from the Fastest Q&a Site in the West. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2857–2866). New York, NY, USA: ACM. http://doi.org/10.1145/1978942.1979366

Pal, A., Chang, S., & Konstan, J. A. (2012). Evolution of Experts in Question Answering Communities. In *Sixth International AAAI Conference on Weblogs and Social Media*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4653

Pal, A., & Konstan, J. A. (2010). Expert Identification in Community Question Answering: Exploring Question Selection Bias. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1505–1508). New York, NY, USA: ACM. http://doi.org/10.1145/1871437.1871658

Schneider, J., Passant, A., & Decker, S. (2012). Deletion Discussions in Wikipedia: Decision Factors and Outcomes. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (pp. 17:1–17:10). New York, NY, USA: ACM. http://doi.org/10.1145/2462932.2462955

Tausczik, Y. R., Kittur, A., & Kraut, R. E. (2014). Collaborative Problem Solving: A Study of MathOverflow. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 355–367). New York, NY, USA: ACM. http://doi.org/10.1145/2531602.2531690

Tausczik, Y. R., & Pennebaker, J. W. (2012). Participation in an Online Mathematics Community: Differentiating Motivations to Add. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 207–216). New York, NY, USA: ACM. http://doi.org/10.1145/2145204.2145237

Treude, C., Barzilay, O., & Storey, M.-A. (2011). How Do Programmers Ask and Answer Questions on the Web? (NIER Track). In *Proceedings of the 33rd International Conference on Software Engineering* (pp. 804–807). New York, NY, USA: ACM. http://doi.org/10.1145/1985793.1985907

Wohn, D. Y. (2015). The Effects of Feedback and Habit on Content Posting in an Online Community. Retrieved from https://www.ideals.illinois.edu/handle/2142/73646

Yang, J., Hauff, C., Bozzon, A., & Houben, G.-J. (2014). Asking the Right Question in Collaborative Q&a Systems. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (pp. 179–189). New York, NY, USA: ACM. http://doi.org/10.1145/2631775.2631809

Zhu, H., Kraut, R. E., Wang, Y.-C., & Kittur, A. (2011). Identifying Shared Leadership in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3431–3434). New York, NY, USA: ACM. http://doi.org/10.1145/1978942.1979453

Zhu, H., Kraut, R., & Kittur, A. (2012). Effectiveness of Shared Leadership in Online Communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 407–416). New York, NY, USA: ACM. http://doi.org/10.1145/2145204.2145269

Zhu, H., Zhang, A., He, J., Kraut, R. E., & Kittur, A. (2013). Effects of Peer Feedback on Contribution: A Field Experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2253–2262). New York, NY, USA: ACM. http://doi.org/10.1145/2470654.2481311