

When Interaction is Valuable: Feedback, Churn and Survival on Community Question and Answer Sites: The case of Stack Exchange

Benny Bornfeld
School of Engineering,
Ruppin Academic Center,
Israel
and Center of Internet Research,
Faculty of Management,
University of Haifa,
Israel
bennyb@ruppin.ac.il

Sheizaf Rafaeli
Center for Internet Research,
University of Haifa,
and Samuel Neaman Institute,
Technion,
Israel
Sheizaf@rafaeli.net

Abstract

Does feedback contribute to collaboration? As in most open participation and contribution platforms, churn is an issue. The highest churn and dropout rates follow the initial posting of a single answer. According to feedback theories, contributors are sensitive to feedback. Votes and comments are common feedback mechanisms in such platforms. Prior studies on the effect of these mechanisms in different platforms have produced conflicting results.

This study reports a longitudinal analysis of the feedback effect on newcomer answer provider retention in five Stack Exchange communities, including over a million users and their answers. We find that feedback in the form of votes and comments provided to the first answer is strongly correlated with newcomer retention. Thus, interaction is valuable. The findings have implications for the design of Q&A websites and for testing the theory of feedback arrangements' impact on persistence.

1. Introduction

The survival of open community content platforms depends on continued contributions by their members. Churn, i.e. the desertion of contributors, is a major concern in many online platforms[1,2,3,4]. The highest churn rate appears after a contributor's single (first and last) contribution.

Feedback is known to affect churn[5], especially of newcomers[6]. In this study, we focus on the role of feedback mechanisms on preserving contributors

beyond the first answer in community question-answering (CQA) websites. Specifically, we focus on the roles of vote and comment feedback mechanisms.

This issue has been studied by other scholars and there are some contradicting reports as to the effect of the various mechanisms involved (Table 1). One of the main goals of this study is to provide evidence that will shed light on these contradictory findings.

In order to study the feedback effect on contribution survival, we gathered and analyzed data from five Stack Exchange communities. Stack Exchange is one of the world's largest and most successful CQA services. Recently, answer-question ratio on Stack Exchange has been decreasing[7], suggesting that the survival challenge is mainly in maintaining answer providers. Hence, in this paper, we examine the effect of vote and comment feedback mechanisms on the survival of answer providers only.

The feedback process involves two parties: providers and receivers. The platform designers have made several attempts to influence feedback, such as encouraging voting, especially on newcomers' posts. An example is the "Summer of Love" – a call made in 2012 for more positive feedback on newcomers' contributions. In order to identify interventions and changes of this nature over time, we analyzed all data from the nine years in which the service operates. Our analysis shows a strong effect of both votes and comments on the persistence of first time answer providers.

The rest of the paper is organized as follows. We start by presenting the churn issue and follow by outlining the main relevant feedback theories. After reviewing the literature, we present our hypotheses on

the effect of the two feedback mechanisms on contributor survival. Following the Method section, we analyze and discuss the results. We conclude with presenting the implications and limitations of our findings.

2. Churn analysis

Churn is a major issue in all online communities. The highest desertion rate is witnessed following a single contribution. Sixty-eight percent of newcomers to Usenet groups have never been seen after their first post[10], 54 percent of the developers in the Perl open-source development project have never returned after posting a single message[11], 60 percent of registered editors in Wikipedia never make another edit after their first 24 hours[12].

Measuring churn in services that are free of charge is somewhat tricky. Users do not announce their desertion – they just stop contributing. While there is always a chance that users would return to contribute, as time goes by, it diminishes. In order to verify this behavior for Stack Overflow, we examined the interval between the first two answers users provided. About half posted the second post within a month, 84% within one year and 93% within two. Therefore, when analyzing churn rates, we examined only one year from the last post in our dataset.

Figure 1 shows the churn for answer providers per the number of answers. The three color bars represent data from the three largest Stack Exchange websites: Stack Overflow, Super User, and Mathematics. As in the other communities described above, about half the users desert after posting a single answer.

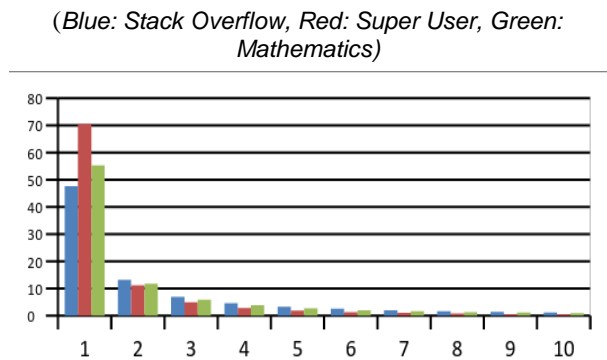


Figure 1: Percent of answerers' churn by the number of the answers posted.

Why do users stop contributing? Understanding churn requires us to explain contribution first. The three prerequisites for users' active participation are expertise, time and most importantly, motivation. The literature describes different motivations which fall

under six categories: getting information, giving information, reputation building, relationship development, recreation, and self-discovery[13,14,15]. These motivations are reflected in Stack Exchange's internal annual survey.¹

Contribution motivation categories such as reputation building, relationship building, and self-discovery are regulated by motivational affordance mechanisms such as points[8,16], badges[17] and feedback[10,18]. In this study, we hope to contribute to the understanding of the influence of feedback on newcomer motivation to continue contributing to CQA services.

3. Feedback theories

Feedback is one of the most powerful influences on learning and achievement, but this impact can be either positive or negative.[19]

Kluger and DeNisi[5] define feedback intervention as: "actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one's task performance". Feedback theories are at the heart of behavioral psychology. Thorndike's Law of Effect[20] states that behaviors followed by satisfying consequences tend to be repeated and vice versa. Feedback interventions strongly influence both pleasantness[21] and arousal[22]. Hence, positive feedback that leads to pleasantness and arousal would lead to repetition – more content contribution in our case. The result of negative feedback is less clear. Arousal and unpleasantness have contradictory effect on further contribution.

In contrast to the parsimonious nature of Thorndike's Law of Effect, Kluger and DeNisi[5] presented Feedback Intervention Theory (FIT). FIT emerged following evidence that many observations were inconsistent with the Law of Effect. FIT proposed several feedback-related constructs, including hierarchy and limited attention. Hierarchy refers to the level at which feedback is perceived. Although feedback is given at the task-level, the receiver can relate the feedback to the self-level, viewing it as personal feedback. FIT assumes that "attention is limited and therefore only feedback-standard gaps that receive attention actively participate in behavior regulation". The first feedback a newcomer receives warrants high attention[1,23].

¹<https://insights.stackoverflow.com/survey/2017>

4. Related work

Several scholars have studied the effect of feedback mechanisms on contribution and churn. This section

Table 1. Summary of related work

| Main Findings | Platform | Paper |
|---|--|---------------------------|
| Answers' positive effect: weak support for best answer (YA, BK); partial support for points and comments (BK) | Baidu Knows (BK), Yahoo! Answers (YA), Naver Knowledge-iN (NK) | Yang et al. [18] |
| Moderation feedback & replies affect participation. Newcomers abandon if they receive no attention. No comments lead to 12% increase in desertion. No difference between up and down votes on first message | Slashdot – news & discussion website | Lampe & Johnston [25] |
| Churn predictors: time between posts, answering speed, reputation of answerers, no. of answers. No comment effect | Stack Overflow | Pudipeddi et al. [2] |
| Comments have a positive effect on the value of answers | Google Answers | Raban [29] |
| Reputation (points) increases participation. Comments decrease participation | Math Overflow | Tausczik & Pennebaker [8] |
| Sharing: for those who are inclined to contribute, receiving feedback increased sharing | Facebook | Burke et al. [26] |
| Positive effect: response to the first post. The parameters of actually getting a right answer or emotional tone have no effect | Six newsgroups | Joyce & Kraut [10] |
| People who are highly sensitive about what others think of them and have high self-esteem are more likely to perceive higher social support from PDAs | Facebook | Whon et al. [27] |
| Task level increases with comments. Ego level increases with grades and praise | 5 th & 6 th -grade students | Butler [24] |
| Reverts demotivate, lead to high churn but more quality work | Wikipedia | Halfaker et al. [1] |
| Reasons for newcomer churn: higher percentage of being deleted and getting no answer | Stack Overflow | Slag et al. [23] |
| Strong peer feedback effect on newcomers. Negative and directive feedback increase task effort. Positive feedback decreases it. Positive and social feedback increase general motivation. Negative feedback decreases general motivation. | Wikipedia | Zhu et al. [23] |

presents work done on different CQA as well as other platforms such as Wikipedia, newsgroups, and Facebook. We have organized these studies in three strands: those dealing directly with feedback and activity lifespan, those dealing with feedback and contribution; and those that focus on the role of comments. Overall, we find that different studies reported significantly different results (see Table 1 for a summary).

4.1 Feedback and activity lifespan

Yang et al.[18] studied three CQA platforms: Baidu Knows (BK), Yahoo! Answers (YA) and Naver Knowledge-iN (NK). They reported the following with regard to predicting the activity lifespan by the first answer: "Consistently between YA and BK, having one's answer selected to be the best is a promising sign for a longer lifespan. On BK, earning points also had a positive effect, and importantly, getting feedback about the answers from the asker (best Commented) also was correlated with users staying longer".

Halfaker et al.[1] studied the effect of reverts in Wikipedia on newcomer retention. They reported a dramatic drop from 40% before 2005 to 12-15% after 2007 for returning newcomers. They concluded: "Reverts are powerfully demotivating, but their net influence is that more quality work is done on Wikipedia as a result of reverts than is lost by chasing editors away".

Joyce and Kraut[10] studied retention in six newsgroups. They found that contributors' probability of re-posting increased from 44% to 56% after receiving a reply to their initial post. They reported being surprised to find that the quality of the response—its emotional tone and whether it answered a newcomer's question—did not influence the likelihood of re-posting.

Finally, Pudipeddi et al.[2] studied newcomers and veterans' churn characteristics and predictors in Stack Overflow. They reported that "the time gap between subsequent posts is the most significant indicator of diminishing interest of users, besides other indicative factors like answering speed, reputation of those who answer their questions, and number of answers received by the user." Their model included comments but they did not report that they had had any effect on churn.

4.2 Feedback and contribution

Zhu et al.[23] examined the effect of peer feedback on contribution in Wikipedia based on FIT. They manipulated users by sending feedback messages of

four types: positive, negative, social and directive, and found a strong effect on newcomers. They distinguished between efforts on a specific task and general motivation. In accordance with FIT, they found that on the task level, both negative and directive feedback had a strong effect on increasing the effort while positive feedback decreased effort. Regarding general motivation, they found that positive and social feedback had a strong positive effect while negative feedback has a negative effect.

Butler[24] studied different feedback interventions in fifth and six-grade classes. She reported that effort, outcome, and impact on evaluation of task-involved causes were highest after receipt of comments. Ego-involved attributions were highest after receipt of grades and praise.

Lampe and Johnston[25] studied the effect of feedback on participation in Slashdot, a news and discussion site. They reported that moderation feedback and replies affected participation and that if new members received no attention from the community, they were likely to desert, feeling they were not appreciated. According to their data, there was a 12% increase in desertion in case of no feedback, but interestingly no difference between up-voted and down-voted first-time messages.

Tausczik and Pennebaker[8] studied the motivations for participation in Stack Exchange math community. They argued that positive voting feedback encouraged further posting. On the other hand, they found comments to be demotivating: "Unexpected findings were found for community responses related to constructive feedback. Receiving comments discouraged participation in two of the models. When comments were given, three models showed that disagreement significantly encouraged participation and one showed it was marginally related to participation. Agreement in comments was not related to participation."

Burke et al.[26] studied newcomers' contributions on Facebook and reported that "for newcomers who are initially inclined to contribute, receiving feedback and having a wide audience are also predictors of increased sharing". In another study on Facebook, Whon et al. [27] explored the effectiveness of "likes" and concluded: "People who are highly sensitive about what others think of them and have high self-esteem are more likely to perceive higher social support from paralinguistic digital affordances[PDA](e.g. likes)".

4.3 The role of comments

Vargo and Matsubara[28] studied the role of comments on low-quality questions in Stack Overflow. They found that the most popular and frequent

comments included criticism, which was not aligned with the declared norms of the service. Anderson et al.[17] reported that in Stack Overflow, "Community interaction in the form of comments on answers has a significant predictive power on the long-lasting value of a question". In a study on the paid service of Google Answers, Raban[29] reported that satisfaction was improved when answerers also provided free comments. Conversely, Ahn et al.[30] examined how users in several Stack Exchange communities learned to be better askers and found no correlations with comments received on previous questions.

4.4 Related work summary

Prior studies have examined the effect of feedback, votes and comments on users' retention and contribution. There are several inconsistent findings. The main findings are summarized in Table 1. The main inconsistent studies per mechanism are:

Points: Tausczik and Pennebaker[8] report that points increase participation while Lampe and Johnston[25] find no effect and Yang et al.[18] reports partial support.

Comments: Tausczik and Pennebaker[8] find that comments decrease participation, Pudipeddi et al.[2] do not report an effect, Yang et al.[18] find partial support and Lampe and Johnston[25] report a strong positive effect on participation and value.

In light of these inconsistencies, we examine the direction of the feedback effect on survival and its magnitude.

5. Research questions

Our goal is to measure the effect of feedback mechanisms on the persistence of newcomer answer providers in community question-answering services. The independent variables in explaining newcomers' survival are the answer's *score* (i.e. aggregated votes) and the existence of *comments* on the answer. We operationalized score into four categories: Accepted, Positive, Zero and Negative. Accepted votes are awarded by the asker to only one answer which he regards as most useful. We operationalized comments into two categories: With Comments and Without Comments. The rationale for this operationalization is described in the Method section. The dependent variable, *survival*, is dichotomous. It is true in case the user has posted more than one answer and false in case she has not.

Following Thorndike's Law of Effect, we expect that a Positive score and Accepted vote feedback will have a positive effect on contribution survival. Given

that newcomers are more sensitive to feedback, we expect a strong effect.

H1: The order of the effect size of *vote categories* on contribution survival would be Accepted > Positive > Zero > Negative.

H2: *Comments* have a positive effect on contribution survival.

We expect that the net effect of a second feedback mechanism would be lower than in the case of a single mechanism or a stronger signal:

H3: The effect size of comments for the different categories would be Negative > Zero > Positive > Accepted

H4: The effect size of vote categories would be higher for answers without comments.

6. Method

The dataset was derived from the five largest Stack Exchange communities: Stack Overflow, Super User, Mathematics, Ask Ubuntu and Server Fault (Table 2). Due to its relative size and extensive references in related work, we first analyzed the behavior for Stack Overflow and then checked whether the findings were consistent in the other communities. The data were queried using the TSQL interface from the Stack Exchange data website.²

6.1 Stack Exchange

The Stack Exchange service, home to over 150 active communities, is one of the world's leading CQA platforms. Its first and largest site, Stack Overflow, was launched in 2008. Community topics range from the technical to hobbies and other areas of life (e.g. programming, cooking, languages, and parenting). Community sizes range from thousands to millions in both numbers of posts and users.

Stack Exchange is an open service. Anyone can register and contribute to it. Its Q&A repository is open to all without a need to register and most of its users are not registered. Registered users participate in asking questions, providing answers, voting on the usefulness of posts, commenting and editing other's posts. Voting, commenting and editing rights are limited to users above a certain reputation score. High reputation users also undertake moderating roles such as closing and deleting improper or low-quality posts.

² data.stackexchange.com

Stack Exchange employs a wide set of motivational gaming mechanisms: points (i.e. reputation), badges, leaderboards, bounties and secret hats. Reputation is a major motivation for participation in CQA[8,9]. Users gain reputation if their posts are accepted or up-voted by others, and lose reputation when their posts are down-voted or deleted (Table 3). Table 2 presents descriptive statistics on vote and comments in the five communities. In general, there is similarity in the votes and comments' category distribution for first-time answerers across the different communities.

Table 2. First answers, score type and comments

| | #first Ans wers | % Acce pted | % Posit ive | % Zero | % Nega tive | % With Comm ents |
|-------------------|-----------------------|-------------------|-------------------|-----------|-------------------|---------------------------|
| Stack Overflow | 854 K | 21 | 32 | 43 | 5 | 33 |
| Super User | 57K | 17 | 39 | 38 | 6 | 38 |
| Mathematic s | 32K | 20 | 41 | 33 | 6 | 38 |
| Ask Ubuntu | 47K | 16 | 40 | 39 | 5 | 37 |
| Server Fault | 27K | 22 | 40 | 32 | 6 | 33 |

Table 3. Stack Exchange voting mechanism

| Reputation change | Type |
|--|----------|
| +10 for the answer provider | Upvote |
| -2 for the answer provider, -1 for the voter | Downvote |
| +15 for the answer provider | Accepted |

6.2 Design considerations

In this study, we examined the churn of newcomers who provided answers. For simplicity, and because question asking and answering were driven by different motivations, we excluded asking behaviors from the analysis. Since activity history may play a role, this design consideration placed a limitation on this study.

As discussed in the section on churn, most second posts occurred in the first year (84%). Hence, the records of last year in the dataset were excluded from the analysis.

6.3 Data limitation

Stack Exchange removes user identity from deleted posts. Hence, we are not able to measure the survival of users whose first post was deleted. In Stack Overflow, 15.2% of the answers were deleted. The percentage of deleted posts among posts with a negative score was high (~70% in Stack Overflow) and there was also a considerable amount of zero score

posts which got deleted (~22% in Stack Overflow). All the churn data analysis in this study is based on non-deleted posts.

6.4 Analysis over time

Users' behavior in any platform may change overtime because of feature changes, designed behavior manipulation, population or preference changes. Detecting and accounting for such incidents requires analysis over time for all variables. All the longitudinal data we present in this study is aggregated on a monthly level.

6.5 The voting mechanism

Crowd voting is the main underlying mechanism in Stack Exchange. Votes have a direct impact on both post's score and user's reputation. Registered users can vote on post's usefulness (Table 3). A post's score is the number of positive votes minus the number of negative ones.

When a vote is cast, it affects both the post's score and the user's reputation. Hence, the feedback is both on the task level (i.e. post level) and on the self-level (i.e. user level). This is important when examining votes' feedback effect according to FIT. While Stack Exchange stresses that the feedback is on the task-level, the following quote from Stack Overflow meta site exemplifies the hierarchy issue: "Downvotes are exactly for marking problematic answers, @Herr_Doktor; they're not about you. If someone downvoted your answer because it was incorrect, and you got upset about it, that's an adjustment you need to make".³

For simplicity's sake, we classified the answers according to four categories according to their score: Accepted, Positive, Zero and Negative. The Accepted vote is exogenous to score but since it is the best vote one can get we include it as part of the score categories. Accepted answers are categorized under the Accepted category, regardless of their score. Note that Zero scores can result from no vote or from an identical number of positive and negative votes. We checked and saw that over 90% of Zero scores are the result of no votes.

6.6 The commenting mechanism

There are different reasons and motivations for using comments. Comments on answers may provide complementary information (e.g. "A GUID really isn't

necessary") or be expressions of courtesy (e.g. "Thank you for the explanation!") while others aim to clarify the answer (e.g. "where shall I get the Properties?"). Most of the comments communication is between the asker and the answerer. In Stack Overflow, 44% of the comments on answers are by askers, 31% are by answerers, and 25% are left by others.

Stack Exchange instructions state clearly that comments are not for socializing: "If you want to say 'thank you', vote on or accept that person's answer, or simply pay it forward by providing a great answer to someone else's question."⁴

The implementation of the comments mechanism in Stack Exchange is somewhat limited. There are no discussion chains and users can only upvote comments. While Stack Exchange awards gold badges to encourage different activities (e.g. voting, editing), there is no gold badge for commenting.

In this study, we define the presence of comments on answers as a binary independent variable: answers with comments and without comments. The nature of the comment (i.e. gratitude, clarification), the number of comments and the identity of the commentator may have an effect on the research question. This is a limitation of our approach, which can be explored in future research.

In summary, we operationalized the independent variables into a 4×2 matrix of feedback combinations. There are four values in the vote categories (Accepted, Positive, Zero, Negative) and two comment categories (With Comments, Without Comments). The dependent variable is answerers' contribution survival after the first answer.

7. Results

7.1 Voting overtime

Over the years, Stack Exchange designers made several attempts to encourage voting. This included the introduction of three new vote badges and raising the daily voting limit. A specific effort was aimed towards newcomer posts. In 2012, platform cofounder Joel Spolsky acknowledged the issue of newcomer churn and called for a "Summer of Love":

"Newbies will show up, make a newbie mistake... and the old-timers will look at each other... and snort, 'Typical!' ... it will start to feel a little bit unfriendly to outsiders.... This is very dangerous. You have to be able to recruit new members... The success of the community depends on it... The goal is simple: to keep Stack Exchange a welcoming, friendly place without lowering our standards".⁵

³<https://meta.stackoverflow.com/questions/251758/why-is-stack-overflow-so-negative-of-late>

⁴<https://stackoverflow.com/help/someone-answers>

⁵<https://stackoverflow.blog/2012/07/20/kicking-off-the-summer-of->

Figure 3 shows voting pattern over time in Stack Overflow. To observe the patterns of voting on newcomers' answers, it shows the voting newcomers have received (solid lines) and the voting non-newcomers have received (dotted lines). The figure shows that Accepted answer voting for non-newcomers (dotted green) is rather stable (~35%) and so is Negative voting (dotted red) (~3%). As the years go by, Positive voting (dotted blue) declines and Zero scores (dotted purple) increase. Newcomers' receive less Accepted votes (solid green) and more Negative ones (solid red). Examining Positive (solid blue) and Zero (solid purple) voting patterns shows that newcomers have started drifting away from non-newcomers between mid-2010 and the end of 2012. Within this timeframe, there has been an increase in Positive voting on newcomers' answers at the expense of Zero votes. The most evident gap is during the Summer of Love at the end of 2012. It is possible that this deviation is due to intervention by the platform owners, which may have begun prior to their public call for a Summer of Love. We will get back to this anomaly when we analyze the effect of voting on survival.

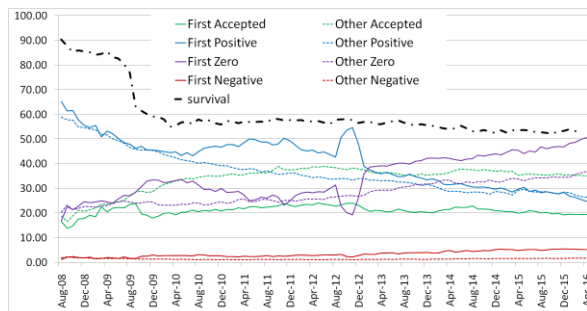


Figure 3: Voting categories' distribution along time in Stack Overflow

7.2 Comments overtime

Figure 4 illustrates the findings of an analysis of the percentage of the posts that have been commented on over time. As seen below, comments on answers were rather stable until 2013, when they start to decline, particularly among newcomers. In the Summer of Love period (i.e. late 2012), there was a local peak in comments on newcomers' answers.

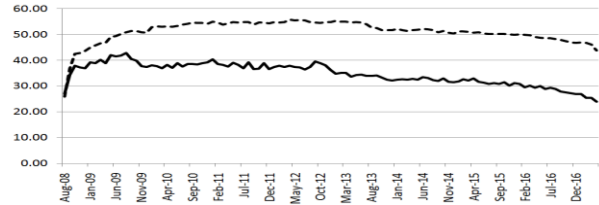


Figure 4: Percentage of answers with comments in Stack Overflow.

Solid – newcomers, dotted – non-newcomers

7.3 Feedback effect on survival over time

Analysis of the relationship between feedback and contribution survival over time presents a complex picture. Figure 5 shows a monthly analysis of survival rates of newcomers in Stack Overflow after posting their first answer. The survival rates are presented according to all eight different feedback combinations. The figure suggests the following observations. Providers of Accepted answers (green) have the highest survival rates and Negatively voted answers (red) have the lowest. Answers with comments (solid lines) are associated with higher survival rates than answers without comments (dotted lines), for all types of votes.

7.4 Behavior inconsistency

Figure 5 shows that Positive and Zero answers produce inconsistent results. In the time frame marked as T3, the survival ratio is lower for Positive (blue) than for Zero (purple), especially for Positive answers with no comments (dotted blue). In T2 and T4, the results are reversed and aligned with the hypotheses. This behavior pattern also appears in the other communities in our dataset. We observe that T3 overlaps with the Positive voting variation gap described above (Figure 3). This anomaly disappears after the Summer of Love. It is plausible to assume that the inconsistency is related to the biased positive voting towards newcomers which peaked during the formal declaration of the Summer of Love.

7.5 The effect of feedback on survival

Due to the anomaly presented in the previous section, the analysis is based on data from T4 (January 2013 until March 2016). The number of observations in T4 is 857K newcomers for Stack Overflow and 163K in all other four communities. The smallest number of observations is for Negative votes with no comments (19K in Stack Overflow and 4K in all the other four communities).

Tables 4.1 and 4.2 present the simple effect of each feedback mechanism, based on monthly aggregated data of all 39 months in T4. Measuring simple effects show the relative contribution of each feedback mechanism. All the figures in these tables are averaged monthly effects (in percent) over the 39 months in the dataset.

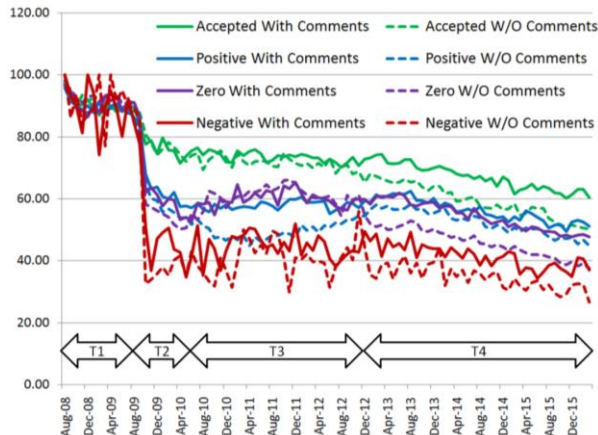


Figure 5: Survival rates in Stack Overflow by different feedback combinations

Table 4.1 shows the effect of vote categories on the survival of answers. We denote the following values: monthly survival rate of answers with Zero score is denoted by $srZv$ and monthly survival rate of answers with Positive score by $srPv$. The effect is calculated as follows: $100 * (srPv - srZv) / srZv$. Negative and Accepted effects are also calculated in the same manner, using the Zero category as baseline. We calculate it separately for answers with and without comments.

Table 4.2 shows the effect of comments on survival. We denote monthly survival ratio with and without comments by $srWc$ and $srWoc$, respectively. The effect is calculated thus: $100 * (srWc - srWoc) / srWoc$.

7.6 Analysis

The results in Table 4.1 provide support for H1. Compared to Zero score that serve as baseline, Positive and Accepted have a positive effect and Negative score have a negative effect. The contribution survival ratio order is Accepted > Positive > Zero > Negative. This holds true for both Stack Overflow and the other communities and both in the presence and in the absence of comments. The vote feedback effect size in the other communities is higher than in Stack Overflow.

Table 4.1. The effect of vote type on answers

| | Negative | Zero * | Positive | Accepted |
|----------------------|----------|--------|----------|----------|
| Stack Overflow (WC) | -23.93 | 0 | 3.27 | 23.07 |
| Others (WC) | -34.30 | 0 | 9.27 | 46.16 |
| Stack Overflow (WOC) | -24.81 | 0 | 14.04 | 29.11 |
| Others (WOC) | -38.24 | 0 | 17.54 | 42.61 |

Averaged monthly effects; WC – with comments, WOC – without comments

() Zero score serves as baseline, $p < 0.001$ (paired t-test).*

Table 4.2. The effect of comments on answers, by the different vote categories

| | Negative | Zero | Positive | Accepted |
|----------------|----------|-------|----------|----------|
| Stack Overflow | 19.76 | 18.37 | 7.19 | 12.83 |
| Others | 30.45 | 21.21 | 11.44 | 23.17 |

Averaged monthly effects; $p < 0.001$ (paired t-test).

Table 4.2 provides support for H2. Comments have a positive effect on contribution survival. This finding is consistent with all vote category types and communities.

H3 is partially supported. Examining the simple effect in Table 4.2 shows that the effect size of comments for the different types is Negative > Zero > Positive, but comments have a strong effect size in the case of Accepted answers: 12.83% in Stack Overflow and 23.17% in other communities. This result is surprising and calls for further analysis.

We have found support for H4. Looking at Table 4.1, comparing the third row to the first and the fourth row to the second yields the simple effect of vote types. In five of six cases, votes have a stronger effect on survival in the absence of comments.

In summary, our results provide support for predictions rooted in Thorndike's Law of Effect. Feedback has an important role in the persistence of newcomers answer providers in community question-answering websites. This finding holds across a variety of types of feedback and across all communities.

8. Summary

Sustaining answer contributors is crucial to the survivability of community question-answering platforms. Related work shows that the highest churn rate is after posting a single answer. Desertion levels after posting a single answer are about 50 percent for Stack Exchange communities. Why do so many users stop contributing after posting a single answer? One possible explanation relates to the role of feedback. Feedback theories postulate that feedback regulates contribution behavior.

The work presented in this paper explores the effect of two feedback mechanisms, votes and comments, on the contribution survival of answer providers after posting their first answer. The effect of these mechanisms was studied by other scholars (Table 1) and their findings were often non-consistent. Providing more evidence to solve these inconsistencies was one of the main motivations for this work.

In order to validate our data reliability, we performed a longitudinal analysis of the patterns of using votes and comments and of the effect of those mechanisms on contribution survival. Our analyses show that between mid-2010 and the end of 2012, there was an intervention, mainly in the voting patterns, which peaked during the so-called "Summer of Love" (Figures 3 and 5). This intervention led to inconsistent effects of feedback on survival behavior. This may explain some of the inconsistencies of studies on Stack Exchange communities during those years. As a result, we did not include this period in our analysis.

In order to validate our findings, we have analyzed and reported the effect on five different Stack Exchange communities. Our main finding is that both votes and comments are strongly correlated with answer contribution survival after the first post. The effect of votes on survival reported here is aligned with Thorndike's theorem, which states that people are encouraged by positive feedback and discouraged by negative feedback[20]. We suggest that the relatively high magnitude of the effect may be related to the construct of level of attention by Kluger and DeNisi's Feedback Intervention Theory [5]. Newcomers are more sensitive to feedback [1, 23]. Negative feedback thus perceived is of course a demotivator.

The positive effect of comments is in contrast to the nil effect reported by Pudipeddi et al. [2] and negative effect reported by Tausczik and Pennebaker[8].

A secondary finding is that the net effect of each feedback mechanism, votes or comments, is stronger in the absence of the other type of feedback.

A somewhat surprising finding is that comments have a strong effect on the survival of answer providers who have received the best type of vote (i.e. accepted answer).

8.1 Limitations

The dataset does not preserve user information for deleted answers. Since most deleted answers have negative scores, the information on negative score answers is incomplete.

Secondly, in this study, comments were treated as a dichotomous variable. As described in section 6.6, comments in Stack Exchange have several roles. Some

provide feedback as to the usefulness of the answer, some express gratification, while others are requests for clarification. Further analysis is needed to distinguish between the comment types and assess their effects.

Finally, this research method cannot claim causality. Our findings show that feedback and contribution survival are correlated. Other, non-controlled parameters may provide alternative explanations for our findings. For example, a-priori motivation to contribute may play an important role in contribution survival. It may have a positive effect on both answer score and contribution survival and explain the relationship between them.

8.2 Future Work

Following the strong effect of comments on contribution survival, future work may explore the effect of different comment types. The surprising magnitude of the effect of comments on answer providers who were given the accepted answer vote calls for further analysis.

We suggest controlling for more parameters such as motivation to contribute. Stack Exchange data contains self-presentation, age and location information that can be used to explore different characteristics that may play a role in the effect of feedback. Future research can examine the relationship between these parameters and sensitivity to feedback.

In this study, we focused on the effect of feedback on answer providers. Future work can analyze the effect of feedback on the survival of contributions by question askers.

8.3 Implications

Stack Exchange introduced three new voting badges[31], increased the voting limit and called for a Summer of Love, which led to an increase in positive votes on newcomers' answers. According to our results, the net effect of positive votes on contribution survival is the smallest of the vote categories.

Comments, on the other hand, seem to be less encouraged in Stack Exchange, as attested to by the absence of a gold badge for commenting. Given the strong effect that comments have on answerers' contribution survival, encouraging comments on answers may lead to higher survival rates.

Acknowledgments

We are grateful to the LINKS Center of Excellence for Learning in a Networked Society and the Center for Internet Research, University of Haifa for support of this study.

References

- [1] A. Halfaker, A. Kittur, and J. Riedl, "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*, 2011, pp. 163–172.
- [2] J. S. Pudipeddi, L. Akoglu, and H. Tong, "User churn in focused question answering sites: Characterizations and prediction," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 469–474.
- [3] S. Rafaeli and R. J. LaRose, "Electronic bulletin boards and 'public goods' explanations of collaborative mass media," *Commun. Res.*, vol. 20, no. 2, 1993, pp. 277–297.
- [4] B. K. Thorn and T. Connolly, "Discretionary data bases: A theory and some experimental findings," *Commun. Res.*, vol. 14, no. 5, 1987, pp. 512–528.
- [5] A. N. Kluger and A. DeNisi, *The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory*. American Psychological Association, 1996.
- [6] R. Kraut, M. Burke, and J. Riedl, "The challenges of dealing with newcomers." MIT Press, Cambridge, MA, 2012.
- [7] I. Srba and M. Bielikova, "Why is stack overflow failing? Preserving sustainability in community question answering," *IEEE Softw.*, vol. 33, no. 4, 2016, pp. 80–89.
- [8] Y. R. Tausczik and J. W. Pennebaker, "Participation in an online mathematics community: Differentiating motivations to add," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 2012, pp. 207–216.
- [9] D. Movshovitz-Attias, Y. Movshovitz-Attias, P. Steenkiste, and C. Faloutsos, "Analysis of the reputation system and user contributions on a question answering website: Stackoverflow," in *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM*, 2013, pp. 886–893.
- [10] E. Joyce and R. E. Kraut, "Predicting continued participation in newsgroups," *Journal of Computer-Mediated Communication*, 11(3), 2006, 723-747.
- [11] N. Ducheneaut, "Socialization in an open source software community: A socio-technical analysis," *Comput. Support. Coop. Work CSCW*, vol. 14, no. 4, 2005, pp. 323–368.
- [12] K. Panciera, A. Halfaker, and L. Terveen, "Wikipedians are born, not made: a study of power editors on Wikipedia". in *Proceedings of the ACM 2009 international conference on Supporting Group Work*, 2009, pp. 51–60.
- [13] U. M. Dholakia, R. P. Bagozzi, and L. K. Pearo, "A social influence model of consumer participation in network- and small-group-based virtual communities," *Int. J. Res. Mark.*, vol. 21, no. 3, 2004, pp. 241–263.
- [14] C. Lampe, R. Wash, A. Velasquez, and E. Ozkaya, "Motivations to participate in online communities," in *proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM, 2010, pp. 1927-1936.
- [15] S. Rafaeli and Y. Ariel, "11 online motivational factors: Incentives for participation and contribution in Wikipedia," *Psychology aspects of cyberspace: Theory, research, applications*, 2008, 243-267. Cambridge University Press, Cambridge, UK
- [16] L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann, "Design lessons from the fastest q&a site in the west," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2011, pp. 2857–2866.
- [17] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 95–106.
- [18] J. Yang, X. Wei, M. S. Ackerman, and L. A. Adamic, "Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities," *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.
- [19] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, 2007, pp. 81–112.
- [20] E. L. Thorndike, "The law of effect," *Am. J. Psychol.*, vol. 39, no. 1/4, 1927, pp. 212–222.
- [21] A. M. Isen, "Positive affect, cognitive processes, and social behavior," *Adv. Exp. Soc. Psychol.*, vol. 20, 1987, pp. 203–253.
- [22] A. N. Kluger, S. Lewinsohn, and J. R. Aiello, "The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal," *Organ. Behav. Hum. Decis. Process.*, vol. 60, no. 2, 1984, pp. 276–299.
- [23] H. Zhu, A. Zhang, J. He, R. E. Kraut, and A. Kittur, "Effects of peer feedback on contribution: A field experiment in Wikipedia," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 2253–2262.
- [24] R. Butler, "Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance," *J. Educ. Psychol.*, vol. 79, no. 4, 1987, pp. 474-482.
- [25] C. Lampe and E. Johnston, "Follow the (slash) dot: effects of feedback on new members in an online community," in *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting Group Work*, 2005, pp. 11–20.
- [26] M. Burke, C. Marlow, and T. Lento, "Feed me: Motivating newcomer contribution in social network sites," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 945–954.
- [27] D. Y. Wohn, C. T. Carr, and R. A. Hayes, "How affective is a 'like'? The effect of paralinguistic digital affordances on perceived social support," *Cyberpsychology Behav. Soc. Netw.*, vol. 19, no. 9, 2016, pp. 562–566.
- [28] A. W. Vargo and S. Matsubara, "Corrective or critical? Commenting on bad questions in Q&A," Mar. 2016. iconferece proceedings
- [29] R. D. Ruth, "Conversation as a source of satisfaction and continuance in a question-and-answer site," *Eur. J. Inf. Syst.*, vol. 21, no. 4, 2012, pp. 427–437.

[30]J. Ahn, B. S. Butler, C. Weng, and S. Webster, "Learning to be a better Q'er in social Q&A sites: Social norms and information artifacts," in *Proceedings of the 76thASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*, 2013, p. 4.

[31]B. Bornfeld and S. Rafaeli, "Gamifying with badges: A big data natural experiment on Stack Exchange," *First Monday*, vol. 22, no. 6, 2017.