AC 2012-3999: STUCK IN THE MIDDLE: THE IMPACT AND PREVALENCE OF FRUSTRATION IN ONLINE QUESTION-ANSWER DISCUSSION THREADS

Mr. Michael Hergenrader, University of Southern California

Michael Hergenrader is a senior majoring in computer science and Spanish. His interests include distributed systems, machine learning, and search technologies. At the Informational Sciences Institute at ISI and at IBM, he is able to work with pride and happiness in all that he does.

Dr. Jihie Kim, University of Southern California

Jihie Kim is the Principal Investigator of the Intelligent Technologies for Teaching and Learning group in the USC Information Sciences Institute (http://ai.isi.edu/pedtek). She is also a Research Assistant Professor in the Computer Science Department at the University Of Southern California (USC). Kim received a Ph.D. from the USC, and master’s and a bachelor’s degrees from the Seoul National University. Her current interests include pedagogical discourse analysis, human-computer interaction, social network assistance, and assessment of student collaborative online activities.

Erin Shaw, University of Southern California

©American Society for Engineering Education, 2012
Stuck in the Middle: The Impact and Prevalence of Frustration in Online Question-Answer Discussion Threads

1. Introduction

The footprint of education in the realm of the internet has dramatically increased over the past decade as university administrators seek to lower costs and still provide classroom education to the maximum amount of people, many of whom would not have access otherwise. To this end, there have even been entire online Computer Science courses released publicly from Stanford and MIT, attracting hundreds of thousands of students from various backgrounds. This transition to the web is not, however, without difficulty or necessary adjustments. With online classes capable of handling ever larger enrollments, instructors are challenged to ensure that each individual’s learning needs are sufficiently met. Asynchronous environments like the online discussion forum are quite a leap from the traditional face-to-face interaction of a classroom.

Thus, it becomes vital for instructors using these tools to understand the needs of participating students and to try to encourage everyone to join in so as to leverage the talents of larger enrollment pools. In this paper, we hope to contribute to the study of these online interactions. Our study views discussion patterns from an emotional point of view, focusing mainly on frustration phrasing in these dialogues. Our aim is to help instructors understand where in student discussions this emotion is found and what factors of the discussion board it affects.

Kim, Shaw, Wyner, Kim, and Li also analyzed student discussions with an emotional approach, but here, we will focus mainly on frustration itself and its effects. We will first examine whether or not the location of frustration in a discussion thread impacts various features of it, including the number of posts in a thread, how many users are actively involved, and the average first response times of posts.

We will then proceed to examine the effects that frustration has with posts close to a deadline. It is intuitive to declare that students normally do not start assigned projects right away; in fact, it is safe to say many do the majority of the work close the deadline. After verifying that frustration often appears in posts made close to a deadline, we will determine whether first responses arrive with more urgency in this situation. From there, we will examine how the length of a frustration-filled question post correlates to first response times and also to response lengths.

2. Study Context

In the context of distributed learning – online education being an example – the frameworks of Communities of Practice and Activity theory have been adopted by practitioners as both aides of understanding and models for development. Wegerif, using a Communities of Practice
framework to analyze accessibility and participation in an online class, found that students’ success or failure in the class depended upon the extent to which students participated online, and whether they belonged or felt like an outsider. Golub elaborated that the mutual exploration, meaning-making, and feedback often lead to better understanding on the part of students, and to the creation of new understandings for them.

Wegerif’s study clearly represented a case for the Sense of Community Theory laid out by McMillan and Chavis, in which members’ needs of learning were met through sharing ideas, successes, and failures. By looking for strategies to promote more effective online forum communication and understanding, we hope to provide a case for this theory as well.

Our study is supported by the National Science Foundation, Research for Experiences for Undergraduates (award # 1045370), CISE IIS (award #0917328), and REESE (award #1008747). A part of the work is supported by the Undergraduate Research Associates Program of the University of Southern California. Our data comes from a challenging university-level operating systems course that offers enrollment for both graduates and undergraduates. These forums provide an aid to students when they are away from the classroom, allowing them the benefits of referencing solutions to similar problems and having their own addressed. Working in teams, students may come in contact with others from their group, those from outside their group, and also with the professor. Eight project forums from two semesters divide up the 418 threads (the whole containing 1841 posts) in our data. Each thread is modeled in a “Q&A” style of discussion, linked up by posts with reply-to relationships. Each project is assigned with specific deadlines of every three to four weeks, depending on the project.

To approach analyzing these threads in an emotional context, we manually annotate general and repeatable English phrases with “emotion acts,” a set of tags similar to those used by Wyner, Shaw, Kim, Li, and Kim, classifying them into different categories that represent common emotions. These tagged phrases are independent of any technical jargon related to this course. However, even with this omission, complications occur frequently in this data set, including misspellings, abbreviations, slang, and inside references. To avoid the bias of one human annotator who may be familiar and tolerant of this noisy content, we employ multiple annotators. To ensure consistency between them, we compare their choices with Kappa scores, a statistical measure of agreement between annotations that corrects for chance agreement. This process was iterative, with annotators comparing results every few threads to ultimately achieve a good measure of above 0.7. Table 1 below describes and provides examples of the “frustration emotion act,” an act used by Wyner, Shaw, Kim, Li, and Kim, that we employ in this study. As seen, there was quite good agreement between the annotators.

<table>
<thead>
<tr>
<th>Emotion Act Tag</th>
<th>Description</th>
<th>Sample Cue Phrases</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRUSTRATION</td>
<td>Expression of hopelessness, individual duress, anxiety, extreme</td>
<td>“I can’t figure it out,” “that sounds pretty tedious,” “HELP,” “not possible here,” “totally weird!”, “I’m</td>
<td>0.72</td>
</tr>
</tbody>
</table>
difficulty; the idea of "going nowhere" is prevalent
badly stuck," “the error still exists,”
“not able to figure it out,” “☺,” “Why is this happening???”

Table 1: Frustration Emotion Act Description

The purpose of the study arose from the predictions about the data mentioned in the previous section: the high amounts of posting activity close to deadlines and the prevalence of frustration-filled posts throughout all four forums, exemplified in Figure 1.

As a sample from one of the two semesters (the other being almost equivalent), the above figure shows the prevalence of frustration throughout the semester and its peaks close to the four deadlines of 9/22, 10/15, 11/2, and 11/21. This figure verifies the earlier notion that students likely haven’t started working on the projects from the start, perhaps enabling more posts to contain frustration due to time constraints. This notion is concretized by the data of Figure 2, enumerating the number of users active each day over the course of the four projects in our 2008 sample.
The local maxima in the graph near each deadline suggest confirmation of the described notion, providing a reason for so much frustration. The questions then arise: what effects does this frustration have on the interactions of these forums? We present the following hypotheses to try and solve this question.

3. “Location, Location, Location”

We first attempt to determine if the amount of frustration phrasing is actually creating more discussion than there would be otherwise. We split the location of frustration in a thread into four categories: no_frus, meaning no frustration shows up in any post; frus_first, indicating that frustration only shows up in the first post; frus_not_first, meaning that frustration shows up in any post except the first; and frus_everywhere, meaning frustration phrasing has been captured in both the first post and afterwards.

The first dependent variable of interest is the average length of a thread for each of these categories – we posited, based on intuition, that more frustration phrasing would yield longer threads. Our hypothesis was partially verified, as seen in Table 2.

<table>
<thead>
<tr>
<th>Thread Subcategory</th>
<th># threads</th>
<th>Average thread length</th>
</tr>
</thead>
<tbody>
<tr>
<td>no_frus</td>
<td>217</td>
<td>3.331797</td>
</tr>
<tr>
<td>frus_first</td>
<td>78</td>
<td>3.089744</td>
</tr>
<tr>
<td>frus_not_first</td>
<td>65</td>
<td>7.030769</td>
</tr>
<tr>
<td>frus_everywhere</td>
<td>58</td>
<td>7.241379</td>
</tr>
</tbody>
</table>

Table 2: Frustration location correlating to the average length of a thread
Our expectations were met regarding the comparison of no frustration in any post to having frustration phrasing appear somewhere in the context of a thread. As seen, the threads with frustration occurring after the initial post – whether the initial one expresses it also or not – are a little more than twice as long. On the other hand, a corollary of our hypothesis was that an initial post that expressed frustration would also be longer. This did not prove correct, as the average thread length of three posts is quite short, nearly identical to the thread length that has no frustration at all.

We then tried to determine correlation between frustration location and another measure of activity: the number of users involved in a thread. We hypothesized that the average number of users in a thread would increase due to frustration, the intuition being that perhaps others would experience the same issue and contribute or have some answer to provide.

<table>
<thead>
<tr>
<th>Thread Subcategory</th>
<th># threads</th>
<th>Average # Users</th>
<th>Average # Student Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>no_frus</td>
<td>217</td>
<td>2.52</td>
<td>1.88</td>
</tr>
<tr>
<td>frus_first</td>
<td>78</td>
<td>2.40</td>
<td>1.92</td>
</tr>
<tr>
<td>frus_not_first</td>
<td>65</td>
<td>3.83</td>
<td>3.14</td>
</tr>
<tr>
<td>frus_everywhere</td>
<td>58</td>
<td>3.66</td>
<td>2.97</td>
</tr>
</tbody>
</table>

*Table 3: Frustration location correlating to number of users involved in a thread discussion*

In Table 3, we present two average counts. The first represents an average over the aggregate set of all users, including instructors, instructor figures, and students. The second reduces the set to only student users. The figures above suggest that our hypothesis was partially correct. Although frustration found in the first post is insufficient to suggest that a longer discussion would occur, frustration found later in the thread confirms that more users become involved in the discussion.

From these results, one could perhaps conclude that the professor or instructor figures tackle and solve the frustration problems quickly. Unfortunately, the disadvantage of this is that other students are not able to attempt a suggestion first. These results support the suggestions of Hergenrader and Kim⁹: in order for instructors to involve more students, they ought to take a passive, supervisory role at first, allowing for others to respond regardless of whether frustration appears.

4. The Wait

But just how long must one wait to receive a response? If the post contains elements of frustration, will these shorten the first response time? Defining a question post time to be at zero with respect to its first response, we choose to address these questions in two stages. First, we measure response times observing just the initial post of all threads and whether it contains frustration. Then, we expand to the set of all question posts. Unlike the earlier hypotheses, these
two sets are not disjoint, as the set of all frustration-filled posts unsurprisingly contains the set of initial ones.

In both stages, we carry out two separate measurements, one for each first response “type” (absolute and relative). We define an “absolute first response” to be the post that addresses its associated question post before any others that might follow it (chronologically speaking). On the other hand, we define a “relative first response” as a post with emphasis placed more on the role of the respondent than on the overall chronological ordering. For this second type, we define two roles, instructor and student, and divide the thread set accordingly to see whether this correlates to how quickly one responds to frustration.

It must be noted that the relative first response is not guaranteed to be absolute. For example, an instructor’s first response to a particular question may or may not be the true first response chronologically speaking – it is just the first for him or her. As a last preliminary measurement, we observe that the average thread lifespan (the time from the first post to the last) in our data set is approximately five days.

We start out simple by testing the absolute first response, determining whether the first response time – regardless of who answered – is altered by frustration. For the initial post, we hypothesize that frustration has a substantial impact on the response time, with the intuition being that the appearance of a new thread containing a frustrated student would evoke a greater sense of urgency among everyone else. Our results appear in Table 4.

<table>
<thead>
<tr>
<th>First post contains frustration?</th>
<th>Number of instances</th>
<th>Average absolute first response time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>130</td>
<td>7 hours, 31 minutes, 29 seconds</td>
</tr>
<tr>
<td>No</td>
<td>275</td>
<td>7 hours, 39 minutes, 50 seconds</td>
</tr>
</tbody>
</table>

Table 4: Average absolute first response times, observing just the first post

Our prediction has been disproven: frustration in the initial post has minimal effect on how quickly the first response will arrive; the eight minute difference is small compared to the magnitude of the seven hours, let alone the average thread lifetime of multiple days.

Still observing the initial post, we turn our attention to the relative first responses for students and instructors, hypothesizing that, with the knowledge of the previous result, frustration in the initial post does not correlate to shorter response times from posts by students or instructors. And since we are now dealing with relative response times, we also make the prediction that student
responses will take longer to arrive than those of the instructor, with the reasoning being that students could feel intimidated due to inexperience with the material.

<table>
<thead>
<tr>
<th>First post contains frustration?</th>
<th>Responses by non-instructor</th>
<th>Average response time (student role)</th>
<th>Responses by instructor</th>
<th>Average response time (instructor role)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>108</td>
<td>9 hours, 9 minutes, 26 seconds</td>
<td>68</td>
<td>15 hours, 16 minutes, 12 seconds</td>
</tr>
<tr>
<td>No</td>
<td>208</td>
<td>10 hours, 29 minutes, 36 seconds</td>
<td>156</td>
<td>17 hours, 56 minutes</td>
</tr>
</tbody>
</table>

*Table 5: Average first response times for the two roles (instructor and non-instructor), observing just the first post*

Table 5 confirms our first hypothesis: while there is some difference between the two times, two hours is a small unit of time when compared with the overall time (weeks) available for a particular project thread and thus negligible. However, it also surprisingly invalidates our second hypothesis: *students*, on average, responded more quickly to the initial question – whether it expressed frustration or not – than instructors, with the former preceding the latter by several hours.

Not wanting to limit our observation to the first question-response pair, we now expand the scope, observing the correlation between frustration and first response times throughout all posts of a thread. Like before, we begin by testing the absolute times, hypothesizing that the result will be the same: that frustration will not shorten the first response time. Our results appear in Table 6.

<table>
<thead>
<tr>
<th>Post contains frustration?</th>
<th>Number of instances</th>
<th>Average absolute first response time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>184</td>
<td>6 hours, 2 minutes, 17 seconds</td>
</tr>
<tr>
<td>No</td>
<td>389</td>
<td>6 hours, 50 minutes, 33 seconds</td>
</tr>
</tbody>
</table>

*Table 6: Average absolute first response times, observing just the first post*

Indeed, as we expand the scope to include all frustration questions and their absolute first responses, the result is the same as before, and our hypothesis is confirmed.

Turning our attention to all relative first responses, we predict that frustration in a question is not enough to shorten the response time. In addition, we hypothesize that instructor responses will arrive later than student ones, keeping in line with the results of the initial post. Table 7 reflects our results.
Post contains frustration? | Responses by non-instructor | Average response time (student role) | Responses by instructor | Average response time (instructor role)
---|---|---|---|---
Yes | 148 | 7 hours, 7 minutes, 37 seconds | 84 | 13 hours, 10 minutes, 9 seconds
No | 300 | 8 hours, 9 minutes, 34 seconds | 186 | 17 hours, 45 minutes

Table 7: Average first response times for the two roles (instructor and non-instructor), observing all posts

Observing all posts, we confirm both hypotheses: the difference in response times is small when compared to the magnitude of a thread’s possible lifetime (the average being multiple days, the maximum being three to four weeks). One to four hours are small units of time in this regard; whether students asking questions were frustrated or not, the response came within a relatively small range of time. Thus, to address the original question of this section, there exists no correlation between frustration phrasing by itself and the time it takes for a first response, whether given by a student or by an instructor, to address it.

In all the previous hypotheses, time has been kept out of the equation – only the thread’s Q&A structure was important. But, as seen in the Study Context section above, there is a pattern that cannot be ignored: a lot more posts are made closer to the project deadlines. Now with regards to time, we wish to know if the deadlines affect the first response times for frustration posts. In effect, does a shorter time window magnify the urgency requested by frustration phrasing, causing responses to come more quickly?

As in the previous section, we split up our data into student and instructor groups. And for each group, we test whether or not frustration in the question post impacts the response times, causing for four total experiments. We hypothesize that frustration phrasing in shorter time windows has a strong impact on first response times: they are shorter for both types of respondents. Overall, we predict that frustration questions will have responses arrive more quickly.

Figure 3: Question post time difference (in seconds) from the deadline vs. student first response times (question has frustration)
Figure 4: Question post time difference (in seconds) from the deadline vs. student first response times (question lacks frustration)

Figure 5: Question post time difference (in seconds) from the deadline vs. instructor first response times (question has frustration)

Figure 6: Question post time difference (in seconds) from the deadline vs. instructor first response times (question lacks frustration)
In the four above figures, each plot point represents a question post. All forum deadlines are represented at x = 0 (this measure of time allows all points to be compared on the same scale). The x-axis itself represents the difference in seconds between the question post time and its corresponding deadline time. On the left side of the y-axis, this difference is negative, meaning that students submitted questions after the deadline but still in the corresponding project forum. The y-axis represents the difference in seconds between the question post time and its corresponding first response time. For all scatter plots, we modeled the data on a continuous basis (since we are dealing with time) through linear regression. This model, based on a least squares error measurement, produced the maximum $R^2$ values and thus was the optimal choice.

However, even with this optimal choice, our hypothesis is proven incorrect. $R^2$ values of approximately zero in all four models demonstrate that throughout the course of a semester (all the slopes of the model lines are essentially zero as well), there is no significant change between the time when a question was asked and when it was answered. In higher level terms, there exists no correlation between first response time and the presence of frustration at different time periods. In addition, the consistent $R^2$ values reveal that the respondent’s role was irrelevant. In earlier sections, we proved that frustration by itself is not sufficient to garner quicker response times. In combination with time remaining until the deadline, this observation holds just the same.

This result, perhaps, can be potentially concerning. As shown, it is true that students tend to start projects later and reach higher levels of discussion board activity near deadlines, but frustration ought to require some sort of urgency, even if it is not close to a deadline. One possible recommendation would be to split project forums by urgency level, advising that those with more concerning problems could post their issues in a separate thread from the others. This could also potentially mitigate the higher activity close to the deadline: if professors require that students post urgent, frustrating problems before a certain point, perhaps the activity around the deadline would not be so hectic.

5. Length

We have seen that limited time is not a factor in having frustration-filled questions addressed more quickly. We also noted the presence of frustration phrasing by itself is also a non-factor. However, this latter experiment can be narrowed down by asking how much language and content does a student includes in a question. If questions are expressed concisely, are students or instructors more likely to answer them quickly? Does frustration play into this? To address these questions, we again have four experiments, splitting the tests and data in the same way as the previous section. We generally hypothesize that a strong correlation exists between the question length and first response time. Our intuition: shorter posts are simpler to address. But more specifically, we predict that a concise frustration post will, unlike previous factors, finally
warrant urgency (and thus shorter response times) while a verbose frustration question will not receive quick responses. In both of these, we believe respondent role will have no effect. We test our hypotheses in the following four plots.

**Figure 7: Question lengths (word count) vs. student first response times (when question has frustration)**

Figure 8: Question lengths (word count) vs. student first response times (when question lacks frustration)

\[
y = 1.426x + 21379 \\
R^2 = 2E-06
\]

\[
y = -5.9536x + 10938 \\
R^2 = 0.0001
\]
As in the previous section, the y-axis indicates the time in seconds between the question and the first response. But this time, the x-axis represents the word count of the post. Again, the linear regression model produced the maximum $R^2$ (correlation) value; thus it was the optimal choice. Although we were correct to predict that respondent role was irrelevant, due to the consistency of the $R^2$ values, our other hypotheses are invalidated. There is no correlation between how wordy the question post is and how quickly a response will come. And regarding our more specific hypothesis, frustration played no role: the response times not differ between frustration and non-frustration questions. The slopes of model lines in all four plots were all essentially flat, dwarfed by the overall magnitude of the y-axis. In fact, the largest slope was approximately a minute and a half, a miniscule amount. Question length is clearly no indicator of how quickly a response will come.
Finally, we test question length (word count) as an indicator of how much the response will say, comparing the first length with the first response length. Does frustration play a part? What about the respondent’s role? We hypothesize that the length of the question post does in fact impact the length of the response: shorter posts will warrant shorter responses, as the problems would seem more basic. However, we hesitate to say frustration posts have more impact or that respondent roles produce any difference; thus, we conclude that there will be no difference from those two facets. Our results appear in the following four figures.

![Figure 11: Question lengths (word count) vs. student response lengths (question has frustration)](image1)

![Figure 12: Question lengths (word count) vs. student response lengths (question lacks frustration)](image2)
The x-axis is the same as in the previous test. The y-axis represents the length of the first response post in terms of words. Our hypothesis is validated to the extent of each of the R-values: there exists a decent correlation between the two lengths in all four cases, given that each is around a 0.20 or above. However, our hesitation to say frustration and respondent role played no part was incorrect, upon consideration of the difference in the first two models. There exists almost a slope of one for when students are responding to frustration-filled questions, whereas the slope of non-frustration questions is only .25. This result suggests that when students answer questions with frustration, they provide more content in their responses. Perhaps they are describing their own experiences with the same problem. Instructors could take advantage of this finding, encouraging students to participate more in threads with frustration posts. By encouraging “crowdsourcing” for a solution to a common problem, an instructor could not only
be free to take a more supervisory role, the idea put forth by Hergenrader and Kim, but he or she could also promote more online participation by highlighting this potential sense of commonality, the idea championed by McMillan and Chavis.

5. Discussion and Future Work

In this study, we looked at post time and thread structure space to try and determine where student frustration fit in and what interaction factors it affected, due to it being the most prevalent emotion in the data set and to the expectations of its appearance during high activity times. We first observed participation variables, in which there did exist a noticeable correlation between thread length and where frustration was located in the thread. However, overall user participation, seen in average user count per thread, did not change significantly. One recommendation for this transition to online space, based on the data from this study, could be that instructors should encourage students to answer these frustration-filled questions first to help establish a more personal (and overall, involved) appeal.

We then observed the impact of frustration on response times. The definitive answer of no correlation came from these four tests, indicating that just because someone was frustrated and desperate for an answer did not mean a quicker solution would arrive.

Finally, we switched to observe a frustration-filled post’s impact with regards to time and length. Neither time to deadline or question length had any correlation with the first response times, frustration or not. However, there was a decent correlation between question and answer length, especially with regards to students responding to their peers’ frustration-laced posts.

To continue the analysis, we would like to first observe other frustration properties in isolation. This includes the “density,” or how often frustration phrasing is found in a post and how it affects participation factors such as those discussed here.

Next, one could conduct tests by combining some of the features. Many of these tests provided strongly negative conclusions regarding the specific feature and the first response time, but perhaps in combination, the results could change. In addition, one could expand the study beyond just the first response, observing the overall timing of responses in a thread.

Finally, given that this subject deals with student learning needs, we would like to observe the effect of frustration on student project performance via grade data. We wish to observe a correlation between how students phrase the question (frustration or not), how quickly responses come, whether the responses provide anything useful (emotionally speaking) instead of just tension-filled answers, and whether longer threads with more students based on frustration helped all involved in contrast to the peers that did not become involved.

2 http://ocw.mit.edu/index.htm


