Superposter Behavior in MOOC Forums

Jonathan Huang  
Stanford

Anirban Dasgupta  
Yahoo! Labs

Arpita Ghosh  
Cornell

Jane Manning  
Stanford

Marc Sanders  
Stanford

ACM’s Learning @ Scale,  
Atlanta, GA  
March 4-5, 2014
2,500s TAs (???)
Benefits of asynchronous online discussions

Scaffolding for learners

- e.g., Wood et al., '76, Zhu, '96

Benefits for motivation, social presence, belonging

- e.g., Rourke et al., '07, Walton et al., '12

Gains in learning outcomes

- e.g., Schellens and Valcke, '06
MOOC forums

Large scale roster, small scale staff => distributed, asynchronous means of communication

Scale is an advantage: chances are better that someone out there will know the answer!
“...a key ingredient in the success of db-class was Amy Cunningham. Amy answered almost 900 questions on the Q&A Forum -- each one patiently, correctly, and at just the right level of detail. I couldn't have done a better job.

Amy Cunningham went on to become a superposter in a number of other MOOCs afterwards...

was a tremendous (some might say indispensable!) resource to tens of thousands of others at the same time.”
The ubiquitous heavy tail

e.g., Wilkinson, ‘08, Wu et al., ‘09
Superposters: The good...

(not to be confused with Superman posters)

Support staff in answering questions
Start conversations
Contribute to “critical mass” of forum participation
Inspire peers by example
Can superposting behavior lead to a “chilling effect” on MOOC forums?
“Someone can't handle the truth.”

“English is a civilized language. Not Hindi.”

“I will predict in advance that the Diamond's analysis of the US will be about 99.9% wrong.”

“You don't know much about history. The US has done more to promote democracy than any other country in the history of the world. Ever. That ain't bragging. It's fact.”
Should we encourage superposters?

What things are characteristic of superposters?

Course designers, pay attention!!

Do superposters suppress forum activity of others?

Hopefully Not.
Anatomy of a thread

Title: Arab Spring; was it all really worth it?

<table>
<thead>
<tr>
<th>Post #1</th>
</tr>
</thead>
<tbody>
<tr>
<td>@user · a month ago</td>
</tr>
<tr>
<td>Two and a half years on from the first spark of the Arab spring, with Egypt still undergoing turmoil, no longer the tourist mecca it used to be, bombings and killings has become a daily news in Cairo. Libya on the verge of breaking into smaller independent states and Syria drowning deeper and deeper into civil war, unable to topple Bashar Al-Assad or even return to their previous peaceful days. Was the Arab spring really worth it after all? Let me hear your thoughts on the topic.</td>
</tr>
<tr>
<td>1 · flag</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>@user · a month ago</td>
</tr>
<tr>
<td>My first thought is that the aftermath of the spring is still running deep, and timely!</td>
</tr>
<tr>
<td>5 · flag</td>
</tr>
</tbody>
</table>

Option to “subscribe” to thread
Comments (optionally) attached to posts
Option to vote (up or down) for post
**Writing in the Sciences**
Non-STEM
~91,000 enrolled
1432 threads
3607 forum participants

**Child Nutrition**
Non-STEM
~32,000 enrolled
1183 threads
1959 forum participants

**Algorithms**
STEM
~44,500 enrolled
1233 threads
2305 forum participants

**Intro to Logic**
STEM
~33,500 enrolled
1021 threads
2665 forum participants

*Caveat Emptor*: Data collected in wild! Draw causal conclusions at your own risk!
Superposters Characterizations

**Quantity Superposters**
Top 5% in course w.r.t. *average number contributions per week*

**Quality Superposters**
Top 5% in course w.r.t. *ratio of user’s # votes to average # votes for other users in thread (averaged over threads)*

**Reputation Superposters**
Top 5% in course w.r.t. *sum of square roots of votes across all contributions*
Superposterhood

Value creation

Relationship with overall forum activity
Superposter Demographics

See also: https://signalblog.stanford.edu/forumdemographics/
Superposters take more courses

![Graph showing the frequency of students registered for different numbers of classes, comparing All Posters, Superposters by Quantity, Superposters by Reputation, and Superposters by Quality. The x-axis represents the number of classes registered, ranging from 1 to 10. The y-axis represents frequency (proportional to # students), ranging from 0.00001 to 1. The graph shows a decreasing trend for all categories, with Superposters by Quantity having the highest frequency at lower numbers of classes and the lowest frequency at higher numbers of classes. All other categories follow a similar trend, with Superposters by Reputation and Superposters by Quality showing slightly higher frequencies compared to All Posters.](image-url)
Cross-course forum behavior

~116,000 total forum participants
~8000 forum participants enrolled in >1 courses
~6200 forum participants enrolled in 2 courses

<table>
<thead>
<tr>
<th></th>
<th>Non-SP in Course #1</th>
<th>Non-SP in Course #2</th>
<th>SP in Course #1</th>
<th>SP in Course #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-SP in Course #1</td>
<td>5328</td>
<td>386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP in Course #1</td>
<td>386</td>
<td>128</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Students > 3X more likely to be SP in both courses than expected if SPness was independent across courses

$\chi^2 = 205.34, p \leq .01$
Superposters outscore peers

Quantity SPs outscored other forum posters by ~1.3 standard deviations

**To control for time of engagement, only considered students who watched > 90% lectures for a course.**
Grade disparity is subject dependent

ML Superposters don’t outperform other forum participants by much

Logic Superposters significantly outperform other forum participants!

Courses that were run more than once

** footnote: baseline is average grade of forum participants, not average grade of all students
Superposterhood

Value creation

Relationship with overall forum activity
“You're in good shape at line 26, but then you go off on a bit of a tangent. For line 27, try assuming \(~(\neg p \mid \neg q)\) instead in order to prove \(~(\neg p \mid \neg q)\) \implies \neg(p & q) for your contradiction…”

“You're really exactly where you want to be Rowena (you just have p and q flipped from the way I did it, which may be why you're not seeing the hint)… …So you can do two assumptions of q and use reiteration… …and the rest of the proof will collapse down to your conclusion in short order.”

**names have been changed to protect the innocent**
“...A little smarter is always better. Learning is always a fun journey.”

“...Thanks to all the people who post the test cases... ...thanks to all the programmer gurus who are willing to help out people in the cloud and make this online course possible as well as enjoyable.”
<table>
<thead>
<tr>
<th>Course</th>
<th># posts from top 3</th>
<th>% on-content posts from top 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child Nutrition</td>
<td>521</td>
<td>81%</td>
</tr>
<tr>
<td>Algorithms</td>
<td>380</td>
<td>80%</td>
</tr>
<tr>
<td>Intro to Logic</td>
<td>551</td>
<td>79%</td>
</tr>
<tr>
<td>Writing in the Sciences</td>
<td>544</td>
<td>39%</td>
</tr>
</tbody>
</table>

**Superposters...**

give more than they receive...
and get fewer downvotes...
and tend to write longer posts...
and respond earlier within threads...

**see stores (paper) for details**

Content from superposters models the behavior we hope to see in any study group, whether online or in-person.
Superposterhood

Value creation

Relationship with overall forum activity
No suppression of activity from others: more posts

** Active students = students who watched > 5% lectures for a course.
No suppression of activity from others: more threads initiated

Non-superposter threads initiated per active student in course

Superposter responses per active student in course

$R^2 = 0.53191$
No suppression of activity from others: more votes received

\[ R^2 = 0.49849 \]

Number of upvotes per active student in course vs. Superposter responses per active student in course.
SPs exhibit high engagement, performance

SP-hood is to some extent an inherent trait

SPs respond faster, write longer posts, are less likely to be down-voted, generally contribute on-content posts

SPs do not suppress activity by other posters
Thank You!