Rating Effects on Social News, Posts, and Comments

Authors: Maria Glenski and Tim Weninger
Presentation: Thomas Kiener, Mike Figliuolo, Bob Reilly

What, Why, How, Constructive Critique

Why:

- Too much information - news, stories, comments...
- Too much content - videos, games, shows, images, music...
- What's valuable and "worth my time?"

Or, from the other side...

"How can I ***make*** people think my content is worth their time?"

To what extent do vote manipulators control the direction of attention and opinion?

Herd Mentality

“The inability or refusal to listen to one's own instinct or 'gut feeling' but to instead follow the majority for fear of being wrong, ostracized or ridiculed.” -Urban Dictionary

Reddit Post Experiment

Sampling:

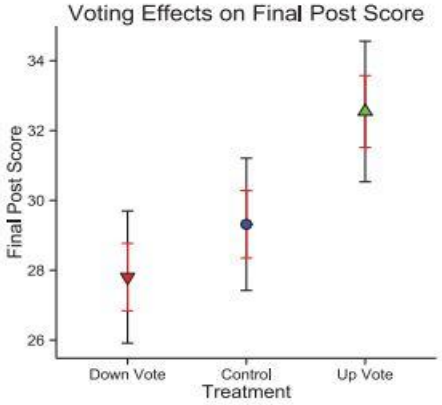
- Posts were sampled over a 6 month period between Sept 1, 2013 and Jan 31, 2014.
- When sampled they were upvoted, downvoted, or left as a control
- Posts were resampled four days later and their final vote totals were recorded.
- 93,019 posts were sampled. 30,998 were upvoted and 30,796 were downvoted.

Reddit Comment Experiment

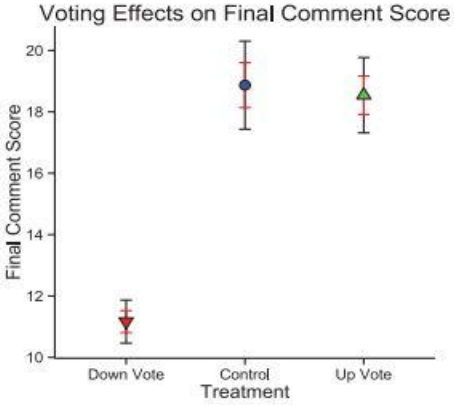
Sampling

- Comments were sampled over a 6 month period between Sept 1, 2013 and Jan 31, 2014.
- When sampled they were upvoted, downvoted, or left as a control
- Comments were resampled four days later and their final vote totals were recorded.
- 96,486 comments were sampled. 35,704 were upvoted and 31,830 were downvoted.

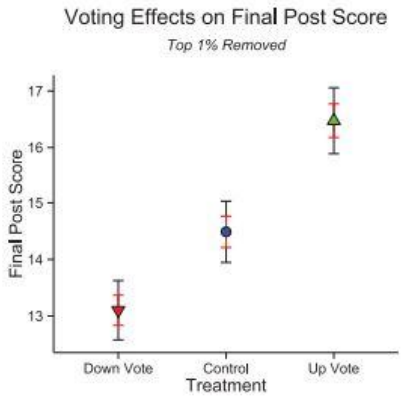
Post and Comment Experiment Results



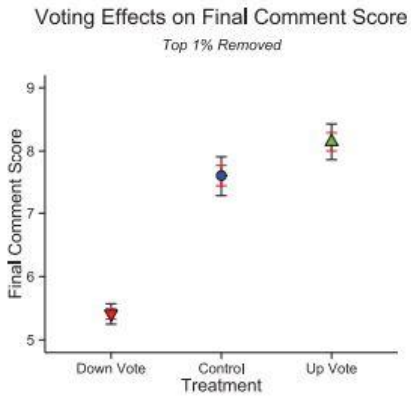
(a) Posts



(b) Comments



(a) Posts



(b) Comments

Black outer error bars show the 95% confidence interval and red inner error bars show the standard error of the mean.

- When including the top 1% of scores, the change in final score is not significantly different from the control group
- This is due to a heavy skew from the top 1%

- To correct for the skewness, the top 1% of scores was removed.
- Tightened confidence intervals.
- Posts that were upvoted and downvoted showed significantly different scores.
- Comments that were downvoted showed significantly different scores.

K-S Test and M-W Test

- The T-test often improperly rejects the null hypothesis when the data is non-normally distributed or highly skewed.
- Use Kolmogorov-Smirnov test and the Mann-Whitney U test.
- Non-parametric tests to compare two unpaired groups of data.
- M-W test ranks all the values from low to high and then computes a p-value that based on the difference between the mean ranks of the groups.
- The K-S test compares the cumulative distribution of the two datasets and computes a p-value based on the largest difference between the two distributions.

K-S and M-W Test Results

Kolmogorov-Smirnov Test:

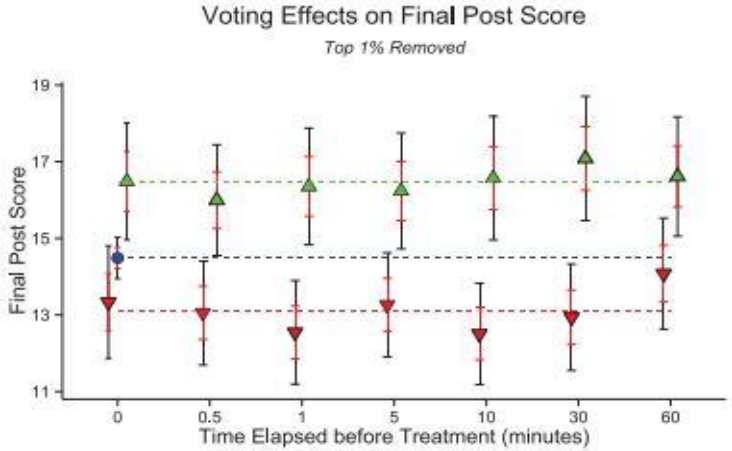
- All p-values came out to be $p < 2.2 \times 10^{-16}$.
- All showed to be statistically significant for differences from the control.

Mann-Whitney U Test:

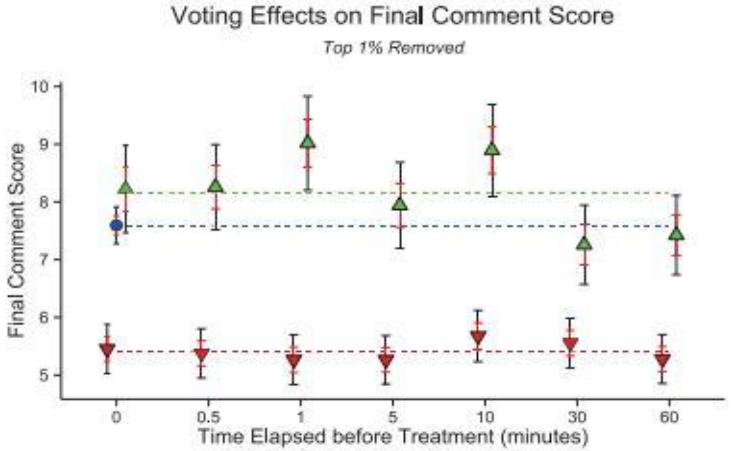
- Upvoted posts $p = 5.9 \times 10^{-53}$
- Downvoted posts $p = 7.8 \times 10^{-73}$
- Upvoted comments $p = 5.57 \times 10^{-15}$
- Downvoted comments $p = 7.52 \times 10^{-8}$
- All are statistically significant for differences from the control rankings.

Does the timing of vote treatment matter?

- Comments and posts are assigned to six random groups.
- Each group is assigned a delay effect. i.e. 0, 1, 5, 10, 30 or 60 minutes.
- It would reason that the early treatment of a post or comment would have a greater impact on the overall score.
- Results for posts did not show a significant impact on overall score.



(a) Posts



(b) Comments

Does the timing of vote treatment matter?

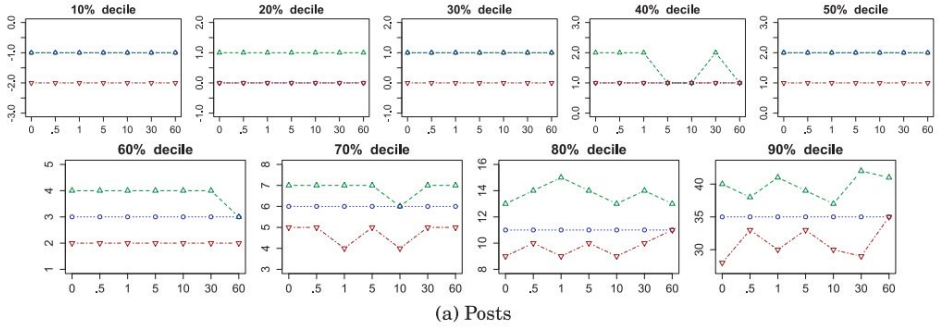
- For posts, displayed error bounds as well as confidence intervals mean very little due to the significant skewing of the data.
- K-S data shows that posts given the upvote treatment were more positively skewed in respects to the control group early on, but as time went on, this diminished.

			0	0.5	1	5	10	30	60
K-S	Post	▲	$D = 0.087$ $p < 2.2\text{e}^{-16}$	$D = 0.087$ $p < 2.2\text{e}^{-16}$	$D = 0.087$ $p < 2.2\text{e}^{-16}$	$D = 0.083$ $p < 2.2\text{e}^{-16}$	$D = 0.082$ $p < 2.2\text{e}^{-16}$	$D = 0.087$ $p < 2.2\text{e}^{-16}$	$D = 0.078$ $p < 2.2\text{e}^{-16}$
		▼	$D = 0.119$ $p < 2.2\text{e}^{-16}$	$D = 0.110$ $p < 2.2\text{e}^{-16}$	$D = 0.110$ $p < 2.2\text{e}^{-16}$	$D = 0.112$ $p < 2.2\text{e}^{-16}$	$D = 0.119$ $p < 2.2\text{e}^{-16}$	$D = 0.097$ $p < 2.2\text{e}^{-16}$	$D = 0.099$ $p < 2.2\text{e}^{-16}$
	Com.	▲	$D = 0.0.43$ $p = 1.8\text{e}^{-07}$	$D = 0.063$ $p = 8.9\text{e}^{-16}$	$D = 0.051$ $p = 3.2\text{e}^{-10}$	$D = 0.045$ $p = 7.0\text{e}^{-08}$	$D = 0.049$ $p = 1.6\text{e}^{-09}$	$D = 0.038$ $p = 1.3\text{e}^{-05}$	$D = 0.021$ $p = 0.04\dagger$
		▼	$D = 0.0.49$ $p = 1.6\text{e}^{-08}$	$D = 0.050$ $p = 9.7\text{e}^{-09}$	$D = 0.049$ $p = 2.1\text{e}^{-08}$	$D = 0.047$ $p = 3.5\text{e}^{-08}$	$D = 0.038$ $p = 2.9\text{e}^{-05}$	$D = 0.031$ $p = 0.1\text{e}^{-02}$	$D = 0.037$ $p = 4.0\text{e}^{-05}$
M-W	Post	▲	$p = 6.1\text{e}^{-14}$ $p = 3.3\text{e}^{-22}$	$p = 1.4\text{e}^{-18}$ $p = 2.5\text{e}^{-17}$	$p = 6.2\text{e}^{-18}$ $p = 4.5\text{e}^{-24}$	$p = 1.8\text{e}^{-13}$ $p = 2.6\text{e}^{-21}$	$p = 2.2\text{e}^{-11}$ $p = 9.9\text{e}^{-27}$	$p = 6.2\text{e}^{-15}$ $p = 4.3\text{e}^{-15}$	$p = 8.6\text{e}^{-12}$ $p = 1.5\text{e}^{-11}$
	Com.	▲	$p = 1.1\text{e}^{-05}$ $p = 0.1\text{e}^{-02}$	$p = 3.7\text{e}^{-10}$ $p = 0.8\text{e}^{-02}$	$p = 1.5\text{e}^{-06}$ $p = 0.3\text{e}^{-03}$	$p = 1.8\text{e}^{-05}$ $p = 3.7\text{e}^{-05}$	$p = 7.3\text{e}^{-09}$ $p = 0.17\dagger$	$p = 3.7\text{e}^{-04}$ $p = 0.31\dagger$	$p = 0.11\dagger$ $p = 0.1\text{e}^{-02}$

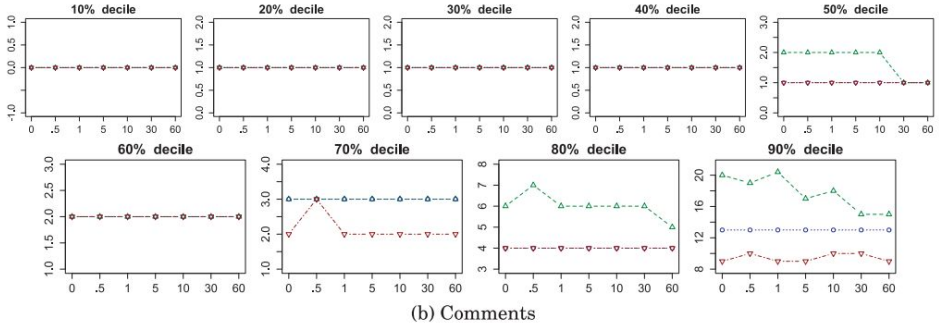
Does the timing of vote treatment matter?

- K-S test results show up-voted treated comments and downvoted treated comments were more skewed than the control group.
- The test statistic value decreases as the delay value increases.
- The timing of the vote treatment plays a greater role in score manipulation for comments rather than posts.
- M-W tests show that post treatment show a significant effect for all delay values with little to no decrease over time.
- M-W tests show a significant decrease of test scores over time, with delays of over 10 minutes shows no significant decrease.

Does the timing of vote treatment matter?



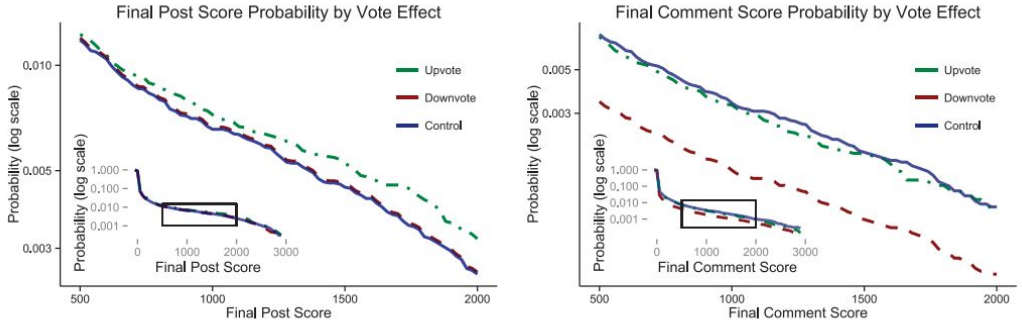
(a) Posts



(b) Comments

How does vote manipulation affect the front page?

- Does vote manipulation affect the probability that a post makes the front page?
- Front page, or trending topic, is defined as any post with a score of 500 or more.



(a) Posts

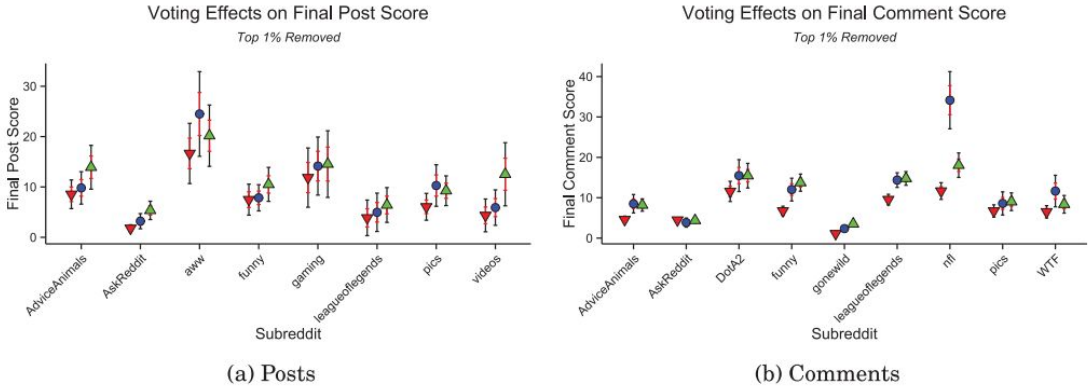
(b) Comments

How does vote manipulation affect the front page?

- Results show that in the upper limit of the distribution, downvote treated posts do not affect the overall score of a post.
- Posts that received the upvote treatment were 7.9% more likely to have a final score of at least 1000 and 24.6% more likely to have a score of 2000 or more.
- For comments, the effect of the upvote treatment or downvote treatment is harder to gauge.
- Comments are inside the post and the hierarchical structure of comment threads hides a significant portion of comments.
- Overall, the upvote treatment had very little effect on the probability of a comment achieving a high score, but the downvote treatment dramatically affected the probability of a comment.

Vote manipulation and subreddits.

- Determined the 10 subreddits that are most frequently in Reddit's rising ranking system as well as subreddits that post content most frequently.
- From these 10 subreddits, two were removed from the group for having a high number of automated posts or having the downvote option disabled.



Vote manipulation and subreddits.

- The M-W test results are similar to the overall results for vote treated posts.
- The M-W test results for comments shows that there are significantly lower scores for the downvoted treated comments.
- The M-W test results show that there is rarely a higher final score for the upvote treated comments.

Conclusions from Results

- Early upvotes/downvotes have strong potential to change how many upvotes/downvotes a post gets in the end.
- An upvote increases a post or comments visibility through an increase in it’s score.
- Posts don’t get many downvotes because they are supposed to be used to mark spam.
- A downvote on a comment has more potential to drop the final score because it causes the comment to be shown far lower in a comment thread.

Critique

- The highest-rated 1% of posts/comments were removed to be able to make conclusions about the results. Those successful posts are the ones we want to learn about the most!
  - Do the paper's conclusions about vote manipulation power hold true for **the most** successful content, i.e. the stuff we actually see on the front page of Reddit?
- Posts and comments with negative scores removed from results.
- Post delay maxed at one hour. Can manipulation occur after much longer time delays?
- Data was only collected once, and conclusions are based on the way the dice rolled this one time. How would results compare across several collection cycles? (Reddit's changes may be to blame for this...)
- Small sample size.
- Too many variables, unfocused. Random subreddit, random up/down/untouched vote, random delay... Tries to ask too many questions with too little data.
- How do we know our randomly chosen posts weren't also manipulated by something even stronger? (Spam/bots/**administrators?**)