Fake News

Harish Tella & Pratik Gami

Research Methods – Final Project

Executive Summary

Facebook reported that the addition of disputed tags on false news stories had the effect of driving more traffic towards these stories (Shu, 2017). We were interested in measuring this effect along with any similar effects from the addition of verified tags to news stories. We sourced two mainstream news articles and two news articles determined to be false. These articles were displayed to survey takers in a format similar to how they would be seen on Facebook's newsfeed. Our survey-takers then rated their likeliness to click on the article. We filtered responses where participants had already seen or heard about an article. After filtering, our study had 90 participants divided into two groups. Each group rated all four articles, but the set of tagged articles was the opposite between the two groups. In the end, we found no statistically significant difference in the likelihood-to-click with the addition of verified or disputed tags.

Problem Context

Fake news has been a huge topic in the cultural consciousness in recent times. The major rise occurred in the run-up to the 2016 U.S. presidential election. It was reported that "1 in 4 Americans visited a fake news website from October 7-November 14, 2016" and that "...Facebook was a key vector of exposure to fake news and the fact-checks of fake news almost never reached its consumers" (Guess, Nyhan, & Reifler, 2018). Origins and motivations behind the creation and distribution of these fake news stories are varied. In addition, the data suggest that most pro-Trump fake news stories reached people who were already heavily favorable to the candidate (Guess, Nyhan, & Reifler, 2018). However, the concerns over the ability to systematically sway public opinion are real.

With the PR damage already dealt and the threat of large scale social manipulation, Facebook sought methods to curb the spread of fake news on their platform. Their strategy was to team up with established groups in the political news media such as Politifact and Snopes. These partner organizations, who already fact-check news articles, would add disputed tags to stories which they considered to have doubtful veracity. Facebook rolled out this process. However, later analysis demonstrated unintended consequences. These disputed tags actually had the opposite effect of their original intention. They actually drove more traffic towards articles that had them. The practice of adding these tags was soon discontinued (Shu, 2017; Clemm, 2018).

Example of story tagged disputed as it appeared on Facebook

This arc of Facebook's foray with tagging, led us to our research interest. If these disputed tags actually increase the traffic towards certain articles, how much does it increase traffic? And then there are alternative approaches. Instead of tagging fake news stories as disputed, Facebook could tag true news stories as verified. How would this approach compare in terms of its effects on traffic?

Related Work

There have been other efforts in curbing the spread of false information on the web that date further back. Ennals, Trushkowsky, and Agosta (2010) built a browser extension that allows users to highlight snippets of text they consider to be disputed. Highlighting text in this manner runs an automated comparison against a database of disputed claims. This database was built on the work of Politifact and Snopes, two notable fact checking organizations.

Users that come across disputed text on a website will see that text as highlighted in red. Clicking the disputed text will brings up a number of related articles with opposing viewpoints similar to the strategy implemented by Facebook recently (Lyons, 2017). There was also an argumentation graph along with other features, but the authors found that users "where more interested in who disagreed with a claim than why people disagreed with it" (Ennals, Trushkowsky, & Agosta, 2010).

Disputed text as it appeared in the work of Ennals et al. (2010)

Although the work of Ennals et al. (2010) stands as a prior work highlighting disputed information on the web, it does not tell us about how their highlighting of disputed text drove movements in user attention. Since these highlights were portions of a web page that the user already navigated to it is more difficult the measure the effect of these highlights. This is compared to the situation on a Facebook newsfeed where we can ask the question “did the disputed tag cause the user to click or not?” Regardless, the work of Ennals et al. (2010) stands as an important contribution.

Now, we would like into introduce a more recent and closely related experiment. Pennycook and Rand (2017) were interested in studying the perceived accuracy of stories that had been tagged as disputed. A large component of this was the implied truth effect. This is an unintended consequence of using disputed tags on news stories. It states that placing disputed tags on false news stories will tend to increase the perceived accuracy of stories that are not tagged disputed whether they happen to be true or false.

In addition to the implied truth effect, the authors were searching for two other effects of disputed tags, the warning effect and the warning backfire effect. The warning effect is the typically expected effect of a disputed tag, where perceived accuracy is reduced because of the tag’s presence. However, the warning backfire effect is when the perceived accuracy of the article is increased because of the presence of the tag. This effect explained by politically motivated reasoning. People may be interpreting the tag’s presence as a sign of bias rather than the process of fact checking (Pennycook and Rand, 2017).

In order to study these research questions, Pennycook and Rand (2017) used an experimental design similar to our own. However, we did not discover their work until we had conducted our own study. Nevertheless, there are a number of differences that are worthy of exploration. To begin, the authors sourced fake news articles from the fact checking website Snopes. They then recruited participants from Amazon Mechanical Turk and divided them up into two groups. One group received 12 real news articles and 12 fake news articles. The other group received 12 real news articles, 6 fake news articles, and 6 fake news articles with disputed tags. Participants were then asked to rate the accuracy of the story and their likeliness to share it.

The author's sample size, N=5,271, allowed them to detect with significance the presence the implied truth effect and the warning effect. However, they did not detect the presence of the warning backfire effect (Pennycook & Rand, 2017). Overall the experimental questions in this study, although not centered around attention shifts, are still critical to study of fake news and its propagation on social media and in the social consciousness. They could form new components of our future studies.

Fake news article with the disputed tag as it appeared in the work of Pennycook and Rand (2017)

Pennycook and Rand (2017) also opted to include politically charged stories among their effect set. There were an equal number of pro-Democrat and pro-Republican articles, and the intensities of political leaning were also balanced. This was achieved with a pre-test where articles were rated on a 5-point scale as pro-Republican or pro-Democrat. Pre-test participants were shown either only fake news articles or only true news articles, but in either case, they were told to consider the articles as true. For the actual experiment, participants had to answer a forced choice question that identified themselves as pro-Hillary or pro-Trump. This allow the data to be appropriately analyzed later. These procedures for handling and testing politically charged material are insightful to us, and could form the basis for future experiments.

Our Study

Now with our context and background in place, we will introduce our study.

<u>Research Question</u>

How much attention do disputed and verified tags draw towards or away from particular news stories? This shift in attention will be measured through users rating their likeliness-to-click after viewing a news story..

Hypotheses

H1: The addition of a disputed tag to a fake news story will increase the average likeliness-to-click versus no disputed tag.

H2: The addition of a verified tag to a real news story will increase the average likeliness-to-click versus no verified tag.

H3: The increase in likeliness-to-click in fake news stories with disputed tags will be greater than the increase in likeliness-to-click in real news stories with verified tags.

Our experimental design

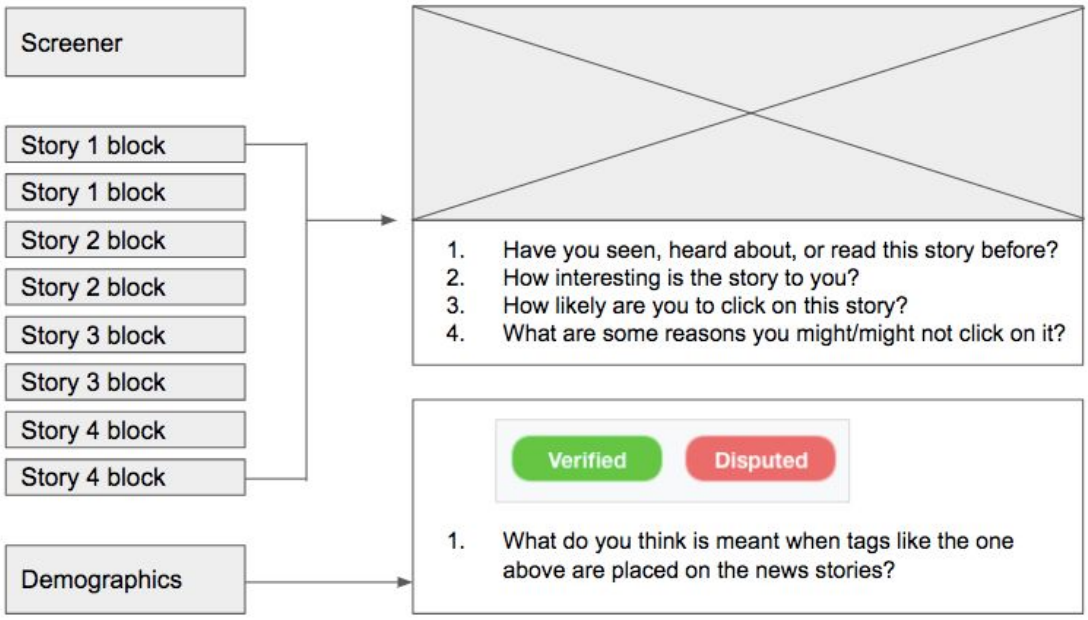
1. Selecting the stories
   We selected two real stories and two fake stories that were published by online news agencies. The criteria for selecting stories were based on the level of controversy, popularity, and believability. For the study, it was important that the stories were not very controversial as controversy can influence the respondents' level of interest and their likelihood to click the story. Also, it was very important for the study that the stories used were unprecedented as there is a higher probability of respondents knowing whether a story is fake or real if the story was known to them earlier. Hence, we stayed away from stories related to politics or sports as these classes of stories are more engaging and had higher chances of being read, heard about, or seen by the respondents. We also stayed away from highly unbelievable stories like UFO sightings or evidence of aliens found on earth, especially while selecting the fake stories, in order to make it less obvious for respondents to conclude on the real or fake nature of the story.

   In another study, when a number of fake news articles were tested with participants, only twenty percent were considered believable (Pennycook, Cannon, & Rand, 2017). We tried to select articles that would fall within this twenty percent.

2. Creating the story presentation format
   After selecting the stories, each story was posted on the researcher's Facebook wall to replicate the visual appearance of a Facebook post. The posts were then edited using the browser debugger. The edits included changing the name of the poster to a generic name in order to keep the researcher's identity confidential. Other Facebook post parameters like time and privacy options were made consistent using the browser debugger. As the final editing step, verified and disputed tags were added to real and fake stories respectively.

3. Survey design

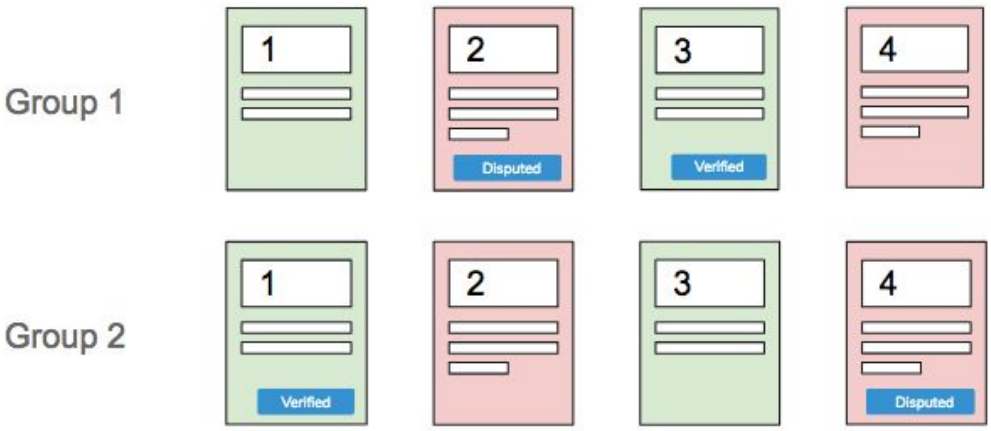
   The survey included three main sessions. The first section included the screener block which dealt with questions related to the recipient's Facebook usage frequency and online news reading frequency. The second section included the story blocks, where each block comprised of a news story and four questions related to the story as shown in the figure below.



   The first question was forced choice with two options 'Yes' and 'No'. The following two questions were rating scale questions and involved a 5-point scale to capture the respondent's level of interest and likelihood to click. The fourth question was an open-ended question intended to capture the reason behind respondent's answer to question three. The final section of the survey included demographic questions and another open-ended question to gain insight into the respondent's understanding of the tags.

4. Survey flow

   The sample was randomly split into two groups. Each group was presented with all four stories. However, a group was either presented the tagged or the untagged version of the stories. For example, if group 1 was presented story 1 without a tag, then the same story was presented to group 2 with a tag. The figure below illustrates this aspect of the survey flow. The order of the story was also randomized for each respondent within a group.



5. Survey dissemination
   The survey was designed using Qualtrics. The link to the survey, along with a brief intro was circulated to:
   - Bentley HFID Certification Facebook Group
   - Bentley HFID Greater Community Facebook Group
   - Bentley HFID California Facebook Group
   - Bentley HFID Slack - General Channel
   - Friends and family
   - Bentley Whatsapp Community Groups
   - LinkedIn posts and groups

6. Sample
   120 respondents answered the survey. The sample, after filtering the respondents (cleaning the data), included 45 males and 45 females and was distributed across seven age ranges as shown in the table below.

15-24	25-34	35-44	35-44	45-54	55-64	65+
13	63	12	7	3	0	1

   The sample recruited from the following groups:

7. Data Cleaning
   The data was cleaned based on the following criteria:
   - Filtered people who didn't complete the survey
     22 people filtered (120 to 98)

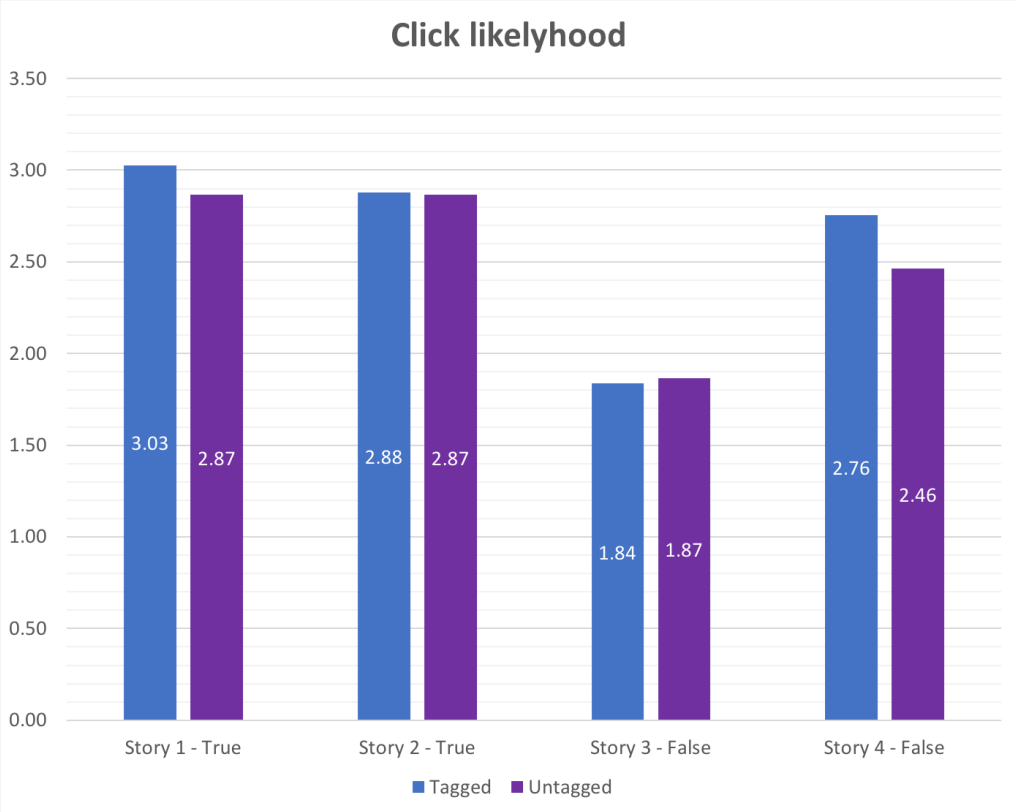
- Filtered people who used facebook once in 6 months or less frequently.
  8 people filtered (98 to 90)

- Filtered people who read news articles once in 6 months or less frequently
  0 people filtered (90 to 90)

- Filter people who answered all 1's all 5's
  0 people filtered (90 to 90)

- Filtered responses to stories that were seen, read or heard before
  8 responses for story 1
  28 responses for story 2
  3 responses for story 3
  11 responses for story 4

8. Results and findings
   The mean rating of the respondents' likelihood to click for each story was calculated. The chart below summarizes the ratings for each story.



The difference between the mean rating of the respondents' likelihood to click on story with a tag and without a tag was calculated in the table below.

		Tag	No Tag	Difference	P-value
Real	Story 1	3.03	2.87	0.6	0.587
	Story 2	2.88	2.87	0.01	0.973
Fake	Story 3	1.84	1.87	-0.03	0.906
	Story 4	2.76	2.46	0.3	0.334

Since the p-values were high (< 0.05), **the data do not support our claims** that 'the addition of a disputed tag to a fake news story will increase the average likeliness-to-click versus no disputed tag' and 'the addition of a verified tag to a real news story will increase the average likeliness-to-click versus no verified tag'.

The mean difference of story 1 (0.6), a real story, is greater than the mean difference of both the fake stories (-0.03 and 0.3). Meanwhile, the mean difference of story 2 (0.01), a real story, is greater than the mean difference of fake story 3 (-0.03) but less than the mean difference of fake story 4 (0.3). Hence, **the results fail to support our third hypothesis**, which is, 'the increase in likeliness-to-click in fake news stories with disputed tags will be greater than the increase in likeliness-to-click in real news stories with verified tags'.

The respondents' level of interest and their likelihood to click on a particular story was found to be strongly correlated. The table below shows the correlation coefficient for all the stories.

	Tagged				Untagged			
	Story 1	Story 2	Story 3	Story 4	Story 1	Story 2	Story 3	Story 4
Correlation	0.90	0.74	0.83	0.78	0.82	0.86	0.85	0.80

Interestingly, 44 People understood the meaning of our tags, 7 People who mention the tags as a reason to click or not click, and 22 People mentioned the source as a reason to click or not click.

Limitations and Future Directions

Limits of Internal Validity

Likeliness-to-click as a proxy for actually clicking

One the major limitations of our study is the survey format. We were unable to actually measure users clicking on things. We were only able to ask users about their likeliness to click. It is unknown how well this serves as a proxy for actually clicking. If it does not stand as a good proxy, then our internal validity is compromised.

We could verify the validity of this likeliness-to-click measure with another experiment where participants receive a stimuli and then they both rate their likeness-to-click and also get the chance to click or not (Krosnick, Judd, & Wittenbrink, 2005). However, with the web scaffolding in-place to conduct such an experiment, it would be easier to simply shift the entire experiment to just checking if users actually click the stories or not.

Tag salience

Our tags are another area of potential concern with regard to internal validity. We had a great deal of freedom in how they were designed and how they were placed on these news articles. It is quite possible they were not salient enough to enter into the survey-taker's conscious awareness. Only seven people mentioned the tags as their reason for clicking or not clicking (note: this open response questions was optional). Increasing the salience of the tags may increase the validity of our experiment and impact our results.

Tag interpretation

A final shadow cast on the internal validity of our experimental instrument was survey-takers' understanding of the meaning of the tags. 44 out of 98 participants responded with the correct meaning of the tags. The rest either did not respond or responded incorrectly. The raises a significant doubt on whether the tags as they stood were sufficient for communicating the idea of 3rd party verification of stories. In hindsight, making this final question on tag understanding option was a mistake. Had we not done this, we could have filtered participants who did not understand the meaning of the tags and still had a significant number of participants.

A Future Solution

All of the previously introduced validity issues could be address in future iterations of this study. We could test the survey with participants following a think-aloud protocol (Fowler Jr., 2013). This would allow us to understand qualitatively how particular questions, survey

elements, and tags are interpreted by participants. This would give us the needed information for iteratively adjusting these elements until the validity is verified.

Tags as they appeared in our survey

Limits of External Validity

Isolated stories

We showed users stories one at a time in order to simplify the mechanics of our experiment and to minimize possible confounding factors. However this has the consequence of limiting the external validity of our experiment. On Facebook, stories appear in a user's news feed. Each element in a user's news feed is competing with the others elements for the user's attention and eventual click. It is likely that tagged news articles have different effects on user attention when they are placed in the context of a news feed, and testing the stories in isolation, as we did in our study, does not capture these effects.

Future Solution

We could extend our study to provide greater generalizability. Although, it would require a significant increase in the complexity of our experimental instrument. We could move away from Qualtrics, and build a custom web application that would allow each story to be presented as part of a news feed or nested within a number of other stories. This would allow the attention competition effects mentioned before to come into play. We leave this as a potential area for future study.

Other Areas for Future Improvement

Other areas for future improvement include increasing our sample size for more statistical power. We could also gain more control over our sample's characteristics. Because of our friends and family recruitment method, we were forced to accept a random collection of participants. It would ideal if we could limit our pool to a more constrained demographic. This would provide us with a greater understanding of our participants' context and give use better basis for interpreting the results.

Another source for future improvement is addressing a possible confound. We found that 21 people mentioned the article's source in their stated reason for clicking or not clicking the article. In this run of the study, we wanted our articles and their presentation format to be authentic as possible. This led us to select articles from real news sources and displaying the articles with the source visible–exactly as they would appear on Facebook. In a future study, we

would need to address the effect of news source on the user's likeliness-to-click. This could be done a number of ways. The easiest method would be to simply omit the news source.

Key-takeaways

Be certain about instrument validity.

We assumed that our participants would understand the meaning of our tags. It is clear that not all of them did. Since much of our data relied on this understanding, we are unsure how much it was compromised. Although we had a question that checked for this understanding of tags, it was optional. The criticality of the question was underestimated.

You need a much larger sample than you think.

Even with 90 participants, our p-values were very high, and we had no significant results. For a study like ours, we need a much larger sample size. This is exemplified by the work of Pennycook and Rand (2017), who in a similar study were able to show statistically significant results. They required over 5,000 participants.

Compromising a bit of external validity is preferred to confounds.

We wanted to have our news articles look exactly like they would on Facebook, but this meant including the article's source. This could have been a possible confound. Next time, we will consider moving away from the exact facebook format for the sake of eliminating that particular confound.

The spread of fake news is a hard problem to solve.

Our background research revealed to us the difficult nature of this problem. Almost every solution seems to have unintended consequences. Nevertheless, the importance of addressing this problem is more clear than ever.

References

Clemm, H. (2018, February 1). Facebook wants its users to drive out fake news. Here's the problem with that. - The Washington Post. Retrieved March 20, 2018, from https://www.washingtonpost.com/news/monkey-cage/wp/2018/02/01/facebook-wants-to-drive-out-fake-news-by-having-users-rate-news-outlets-credibility-heres-the-problem-with-that/?utm_term=.597e901eaa04

Ennals, R., Trushkowsky, B., & Agosta, J. M. (2010). Highlighting disputed claims on the web (pp. 341–350). Presented at the Proceedings of the 19th international conference on World wide web, ACM.

Fowler Jr., F. J. (2013). *Survey research methods*. Sage publications.

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign.

Hughes, T., Smith, J., & Leavitt, A. (2018, April 3). Helping People Better Assess the Stories They See in News Feed | Facebook Newsroom. Retrieved May 4, 2018, from https://newsroom.fb.com/news/2018/04/news-feed-fyi-more-context/

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. *The Handbook of Attitudes*, *21*, 76.

Lyons, T. (2017, December 20). Replacing Disputed Flags with Related Articles | Facebook Newsroom. Retrieved May 4, 2018, from https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/

Pennycook, G., Cannon, T., & Rand, D. (2017). Implausibility and illusory truth: Prior exposure increases perceived accuracy of fake news but has no effect on entirely implausible statements. *Unpublished Paper Manuscript*.

Pennycook, G., & Rand, D. G. (2017). The Implied Truth Effect.

Shu, C. (2017, December 20). Facebook will ditch Disputed Flags on fake news and display links to trustworthy articles instead. Retrieved March 20, 2018, from http://social.techcrunch.com/2017/12/20/facebook-will-ditch-disputed-flags-on-fake-news-and-display-links-to-trustworthy-articles-instead/