Goals in Misinformation

Timothy Kusuma

Faculty Advisor
Stephen Broomell

Abstract

This paper looks at how consumption and supply of fake news are propagated through the intended use of the social media user. We present data from 185 participants from an experiment with participants randomly assigned to generate social media posts with different goals. While we were unable to find a significant result in the area of goal setting, we were able to back up previous work in the field of misinformation.

Introduction

Fake news has become a pervasive topic within the daily lives of those living in the western hemisphere and has become more of an issue as social media has become a bigger part of the way people communicate. While seemingly simple to define as any information that is false or misleading, it is difficult to explain why people fall into the trap of believing in fake news. In this thesis, I present the results of an experiment that seeks to answer a basic question about how fake news propagates. I test whether people are more likely to share unverifiable information when they write a tweet with the goal of being influential.

In the following I present a review of the literature from economics and psychology that discuss the models and social media elements of misinformation. Based on previous work, I pinpoint that goals and incentives involved in social media are a potential mechanism that could increase the probability for individuals to propagate fake news in their tweets. Next I present the results of an experiment designed to test the hypothesis that those who are focused on influencing others on social media are more likely to use false or misleading information. In general, the results failed to support my hypothesis. However, the results of this study demonstrate that people pay more attention to misinformation than the normative economic models would suggest.

Literature Review

Modern economic theory attempts to explain the process of fake news through two theories: Cheap Talk, and Bayesian Modeling. Cheap talk follows a game theoretical model where talk is used as a method for allowing participants to coordinate but falls apart as a desire for coordination disappears (Farrell & Rabin, 1996), hence fake news falls apart as participants no longer find it useful to communicate. One issue with this model is that while there is an incentive to create false information, there is a lack of incentive to believe the information provided. However, we know that misinformation does spread, and people are often susceptible.

Another model, the Bayesian model, takes a more operational approach. This model lays out how different steps in the information distribution process can culminate in allowing people to misinterpret false information as true (Khajehnejad & Hajimirza, 2018). The authors apply this model to understand how reporters deal with information that may not fully link to people's beliefs. The reporters in this model desire minimizing the difference between the original information and the adopted beliefs, while a viewer's adopted beliefs are represented as a linear regression of prior beliefs and new knowledge. This model helps layout the conditions under which misinformation can propagate and what elements help viewers create their beliefs from the information they are presented with.

While these models answer the question of how incentives within the field of misinformation may create and incentivize misinformation spreading, I attempted to adapt this model to investigate how social media posters interact with fake news.

Recent research has shown that people tend to only agree with and share information which is within their homogenous group (Vicario et. al., 2015). So, breaking information into misinformation and scientific news, the longer misinformation propagates the more people who agree with the misinformation, while science news is diffused quickly and reaches its highest impact immediately after publication. Often misinformation that does appear in these homogenous groups will be quickly argued down by members of the group quickly retweeting against fake news (Babcock et. al., 2018).

Other research has shown that while homogeneous political groups are important, analytical thinking plays a major role in whether someone is able to discern whether information is fake or not (Pennycook & Rand, 2018). Building on this analytical thinking approach, other researchers find that goal setting can cause differences in the likelihood of seeking out information while those who are informatively driven to be more likely to ask for information (Ruth, 1993).

I therefore adopt a goal setting framework to theorize about information propagation on social media and take it a step further by testing whether giving social media posters different goals for generating their post would alter their ability to discern which pieces of information were true. I operationalize their ability to discern true pieces of information by observing their endorsement of social media information that is either verifiable or unverifiable. More specifically, I hypothesize that goals of being influential may cause people to unintentionally spread misinformation.

Methods

Participants: We collected a sample of 190 participants in an online mturk study. The participants were told that they would be paid $1.20 for a 10-minute survey where they would read through social media posts and then write their own version. Five participants were taken out of the data, as they provided answers not on the topic of unemployment. The final sample included N = 185 participants, 46.8% female, 26.5% identified as Republican, and 57.8% identified as Democrat.

Design: We employed a between subjects 3 cell design to test the effect of manipulating a participant's goals for writing a tweet on their willingness to cite unverifiable information in their own tweets. Our three conditions were: the control condition, the influential condition, and the informational condition.

To understand how the different goals affected what sources people used when spreading information on social media, we provided participants with a collection of 8 tweets on the topic of unemployment. Unemployment was chosen as it is a topic with differing opinions and evidence is gathered by different organizations that can support two sides of the debate: pro-unemployment benefits and anti-unemployment benefits. As shown in Table 1, these tweets were organized into several groupings, with some being ambiguous where the individual provided an opinion but not any proof to back up their opinion. Others were untrue where the facts presented were intentionally incorrect and the link provided did not connect to any site. And other tweets were backed up by verified information and sent to links that could be considered reputable.

Table: 1 The tweets that participants were presented with. The letters represent the order participants were presented the information.

Tweet Type	Anti-Unemployment	Pro-Unemployment
Ambiguous	A.- Unemployment benefits are hurting our economy. This practice of paying people for no work creates a system of incentives preventing people from going to work, hurting small businesses in the process. These leftist policies need to be stopped. C.- There are more ways to help people than to just throw money at them. Yes, it does help but it needs to be combined with small business loans, and healthcare support. UNEMPLOYMENT is not the only solution.	B.- My view is you can have fairly high unemployment benefits without impacting employment. D.- My wife hasn't received her benefits for 13 weeks. She has to constantly call the unemployment office to just get someone on the line. We have three kids at home and a mortgage that needs to be paid. These benefits would really help. Why hasn't anyone done anything to fix the problem.
Unverifiable	F.- Unemployment is a terrible program. If we look back at the great recession in 2008 we know that the increase in unemployment came from a minimal increase in unemployment benefits leading to 8.3 million more unemployed in 2010 and 6.8 million more unemployed in 2011. Stop this nonsense. Find more here. ● Link not found	E. - The benefits from the Continued Assistance for Unemployed Workers Act only provides $300/week. But it costs on average $500,000 to raise a child, this is not enough money. There is no excuse for not raising unemployment. Find more here. ● Link not found
Verifiable	H.- The unemployment rate has been trending down over the last year. While there seems to have been a strong peak in April 2020 of 14.8 percent this has gone back down to 6.2 percent and looks to continue to fall. So, on the contrary the economy is getting stronger and the need to increase unemployment benefits is overblown. Find out more here. ● tradingeconomics.com/united-states/unemployment-rate#:~:text=Unemployment%20Rate%20in%20the%20United,percent%20in%20May%20of%201953.	G.- A minimum of 10.1 million people are currently unemployed (as of February 2021). There are likely millions of more people struggling with the need of support out in the world. Congress needs to pass stronger unemployment insurance to support all of those struggling with this crisis. Find out more here. ● www.washingtonpost.com/business/2021/02/19/how-many-americans-unemployed/

Procedure: After consenting and reading through the collection of 8 tweets on the topic of unemployment, participants were split into three conditions: control, informational, and influential. In the informational condition participants were informed their goal was to write a tweet that was informative, which we defined as "[allowing] others to make better decisions and [provide] them with the most knowledge." In the influential condition the participants were informed that their goal was to write an influential post, we defined the influential as "one that receives a lot of likes and is retweeted by many others." In the control condition participants were just asked to write a tweet about unemployment. In all of the conditions participants were asked to use the provided tweets in their posts. All posts were limited to 280 character to simulate a social media like experience.

Finally, participants provided demographic information about their Age, Gender, Education, and political leaning and were provided with a debriefing that informed them that some of the tweets we showed them contained incorrect information.

Dependent Variable: We coded each tweet based on whether they draw information from the tweets labeled as "Unverifiable" or "Verifiable" in Table 1. More specifically each individual response was coded on two variables. For the variable labeled verifiable, participant posts were coded with a 1 if their post used the information provided in either of the two verifiable example posts or discussed the topic presented in the verifiable example posts, 0 otherwise. For the variable labeled unverifiable, participant posts were coded with a 1 if their post used the information provided in either of the two unverifiable example posts or discussed the topic presented in the unverifiable example posts, 0 otherwise. Coding was performed blind to condition by the author using a code book that was prepared prior to reviewing the open ended data.

Results

We were interested in understanding how goals affected people's willingness to use information that was unverifiable. More specifically, we wanted to test whether the influential goal of getting more attention on social media increased the proportion of participants that referenced our unverifiable tweets (in Table 1). We plan on testing this by creating two logistic regressions between the information types and conditions.

The overall proportion of participants in each condition using verifiable and unverifiable information is displayed in Figure 1. We predicted that the influential condition should have more people using the unverifiable sources than the control condition. As shown in Figure 1, my goal manipulations appear to reduce the proportion of participants leveraging verifiable sources and increase the proportion of participants leveraging unverifiable sources. However, the effect size in the first place is quite low with most people seemingly not using any of the verifiable or unverifiable information.

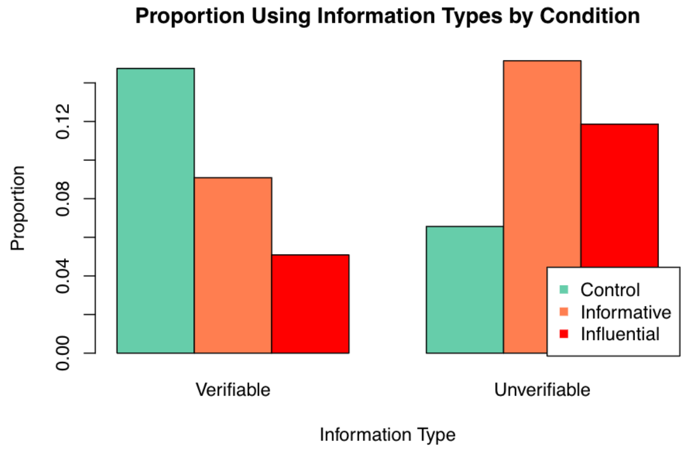

Figure 1: Proportion of participant open ended posts that were coded as using information from a verifiable source (left) and an unverifiable source (right) for each condition.

To see if there was a strong difference between verifiable and unverifiable I ran a logistic regression of the verifiable coded and the unverifiable coded posts and found there not to be a strong significance between the two ($z = -0.832$, $p = 0.431$). I then performed logistic regression on each dependent variable separately given by the following two regression models:

$$\text{Verifiable} = \beta_0 + \beta_1 \text{Influential}_i + \beta_2 \text{Informational}_i \quad \text{Eq. 1}$$
$$\text{Unverifiable} = \beta_0 + \beta_1 \text{Influential}_i + \beta_2 \text{Informational}_i \quad \text{Eq. 2}$$

The influential and informational variables were dummy coded, so each coefficient represents a difference from the control condition.

Table 2 displays the coefficient estimates as beta weights and odds ratios. The regression results reveal no significant effects of condition assignment. Specifically, for predicting the proportion of verifiable sources neither informational goals ($z = -0.549$, $p = 0.327$) or influential goals ($z = -1.173$, $p = 0.091$) significantly differed from control. Similarly for unverifiable sources, neither informational goals ($z = 0.934$, $p = 0.133$) or influential goals ($z = 0.651$, $p = 0.320$) significantly differed from control.

With the limitations due to lack of significance we are unable to conclusively take anything away from these results. However, we can see that the coefficients are moving in the predicted manner where in the influential condition participants are less likely to use the verifiable tweets and more likely to use the unverifiable tweet. While this movement is not significant it does show that that the prediction is moving in the right direction.

In order to increase our power to detect differences, I also ran an exploratory regression model for each dependent variable that included the demographic variables to see if these variables accounted for unmodelled variance in the dependent variables. These results did not change much for predicting the proportion of participants using verifiable tweets. On the other hand, we found a significant effect for gender in predicting use of unverifiable tweets, in that a lower proportion of males tended to rely on unverifiable tweets than females in our sample. Given the exploratory nature, this result would need to be replicated before trusting that it would generalize. However, accounting for this gender difference did increase the effect of the condition assignments (with an increased odds ratio), showing a much stronger effect of information goals in increasing the proportion of participants relying on unverifiable tweets. However, this effect did not reach statistical significance either.

Table 2: Coefficient estimates for logistic regression model for predicting proportion of participants using verifiable and unverifiable sources in their posts.

	Verifiable				Unverifiable			
	Coef (Std)	Odds	Coef (Std)	Odds	Coef (Std)	Odds	Coef (Std)	Odds
Intercept	-1.754* (0.361)	0.173	-0.024 (1.63)	0.977	-2.657* (0.517)	0.070	0.619 (1.805)	1.857
Influential	-1.173 (0.694)	0.310	-2.1 (1.136)	0.122	0.651 (0.656)	1.918	1.092 (1.321)	2.98
Informative	-0.549 (0.560)	0.578	-1.05 (0.814)	0.35	0.934 (0.621)	2.545	2.328 (1.299)	10.254
Gender:Male			-0.679 (0.744)	0.507			-1.931* (0.981)	0.145
Republican			0.702 (1.398)	2.018			-0.572 (1.361)	0.564
Democrat			0.928 (1.167)	2.528			-1.631 (1.171)	0.196
Age			-0.579 (0.442)	0.56			-0.908 (0.511)	0.403

Note: * indicates $p < 0.05$

Conclusion

We wanted to know if goal setting affected the use of misinformation. Taking the results, the lack of significance presents evidence that goals when presenting information on social media has no effect on whether someone is able to filter out data that is true or false. There could be a few reasons for these findings, the first is people are poor at filtering out true or false information. Instead of fake news diffusing as it moves through multiple people it instead moves around like multiple games of telephone as consistent with Vicario et. al. (2015). Then participants would have been more likely to point out issues with the argument they disagreed than with than with the information that was unverifiable (having read through the written posts many articles did seem to fall into this categorization).

Another potential reasoning for these results is a problem with the experiment itself. While we ran our experiment using best practices our process of coding the text may not have been able to pick out all of the elements that the study participants were attempting to signal to us. Our conditions may also have not been clear to those participating in the experiment where we wanted more influential or more informative posts. Instead participants may have ignored this aspect of the question and just focused on writing a post on unemployment.

This research does help add to existing research as it helps show that communication is not cheap talk (Farrell & Rabin, 1996). If it was, participants would treat all of the posts as dubious and refuse to specifically quote any of them. However, the intercepts were significant meaning a significant number of participants were willing to quote a more dubious source than sticking to the more ambiguous sources.

While we were unable to find conclusive evidence there were a few interesting places to take what observations we made from this study. The false or unverifiable tweets were all written using actual numbers and expanding them to seem like better evidence proving that one opinion or another opinion is correct. An interesting use of this element is to use signal detection theory to determine where people stop believing in evidence that supports their world view.

References

Babcock, M., Kumar, S., and Villa Cox, R. A.(2019). Diffusion of pro- and anti-false information tweets: the Black Panther movie case. *Computational and Mathematical Organization Theory, 25*, 72-84. DOI: https://doi.org/10.1007/s10588-018-09286-x

Bessi, A., Caldarelli, G., Petroni, F., Quattrociocchi, W., Scala, A., Stanley, H. E., Vicario, M.D., and Zollo, F. (2016). The spreading of misinformation online. *PNAS, 113*(3), 554-559. https://doi.org/10.1073/pnas.1517441113

Butler, R. (1993). Effects of task- and ego-achievement goals on information seeking during task engagement. *Journal of Personality and Social Psychology, 65*(1), 18-31. DOI: http://dx.doi.org.cmu.idm.oclc.org/10.1037/0022-3514.65.1.18

Farrell, J., & Rabin, M. (1996). Cheap talk. Journal of Economic Perspectives, 10(3), 103–118.

Hajimirza, S., and Khajehnejad, A. (2018). A Bayesian Model for False Information Belief Impact, Optimal Design, and Fake News Containment. *CoRR*. arXiv:1804.01576

Pennycook, G., and Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition, 188*, 39-50. DOI: https://doi.org/10.1016/j.cognition.2018.06.01