Seeing is not believing: Analyzing the Influence of Video-Based Misinformation and Content Flags

Annie Zhu

June 2023

1 Introduction

In the short period of time since social media has become the predominant source of news consumption for the average American, we have already observed a plethora of cases where misinformation has caused real world harm. Notably, there have been examples where false stories have lead to changes in the stock market, potential influences on elections, as well as an increase in vaccine hesitancy, mob lynchings, and other types of widespread panic or criminal activities [1] [2] [3].

Moreover, with the rise of video-based entertainment platforms like TikTok and the dominance of short-form videos in social media content, there is a novel challenge in distinguishing misinformation in online spaces. Current literature on social media and viral misinformation primarily focus on text-based misinformation and image-based content. The current gap in and the importance of understanding effective methods to combat video misinformation motivates my two research questions:

Primary research question: How does the perception and impact of video-based misinformation differ from text-based misinformation? (In other words, does the switch from text to video *cause* an increase in belief in misinformation?)

Secondary research question: When it comes to flagging video disinformation, how do platform nudges and reminders differ in their design and effectiveness compared to text-based misinformation? (In causal terms, do certain types of flags *cause* a decrease in belief in misinformation?)

2 Literature Review

Previous research has shown that flagging of false news on social media platforms like Facebook may indeed help the current efforts to combat sharing of deceiving information on social media [4]. Another study by Lee and Bissell in 2023 explored multiple types of nudges to combat vaccination misinformation and found that both requiring participants to comment and showing participants AI fact-checking labels lead to more positive attitudes toward vaccination [5]. This proposed study contributes to the literature on the effectiveness of content flags by exposing people to two different designs of flags and seeing which is more effective against the control group, which is having no flags.

While text and image-based misinformation have been prevalent, it is crucial to acknowledge that video-based misinformation holds greater potential for misleading the audience and gaining viral traction. Prior literature from 2021 shows that individuals are more likely to believe that an event occurred when it is presented in video form [6]. Sundar et al. indicates that the video format causes news pieces to be perceived as more credible and are more likely to be shared [1]. This proposal contributes to this area of research by testing if video misinformation is more believable than text misinformation.

3 Selection Bias

We need a research design because observing what happens naturally is subject to selection bias on many different accounts. In particular, two clear sources of selection bias is in how social media users act and how social media companies act: First, users select into interacting with video or text based media. It is possible to imagine that younger generations might be more enthusiastic about video and younger generations are less prone to falling for misinformation so as a result it might inaccurately look like video misinformation is less convincing without a research design. Second, platforms selectively choose what types of designs to use to mitigate information. Perhaps platforms prone to having more misinformation on the site could be selecting into strong techniques of mitigation which could make the stronger techniques seem less effective but only because the baseline was different.

4 Research Design

The general approach would be to use a randomized control trial by conducting an online survey using a platform like Qualitrics. The aim would be to achieve a nationally representative sample of 1,500+ individuals aged 18 years and above in the United States.

4.1 Treatment Variables

I propose six different conditions in a 2x3 design, which are visualized in Figure 1. First, I will divide our group of participants up into one subgroup that gets text posts and another subgroup that gets video posts. Then, I further divide each of those two subgroups into three smaller groups for: 1) no flag (control), 2) banner flag (partial obscuration treatment), and 3) blurred screen flag (full obscuration treatment).

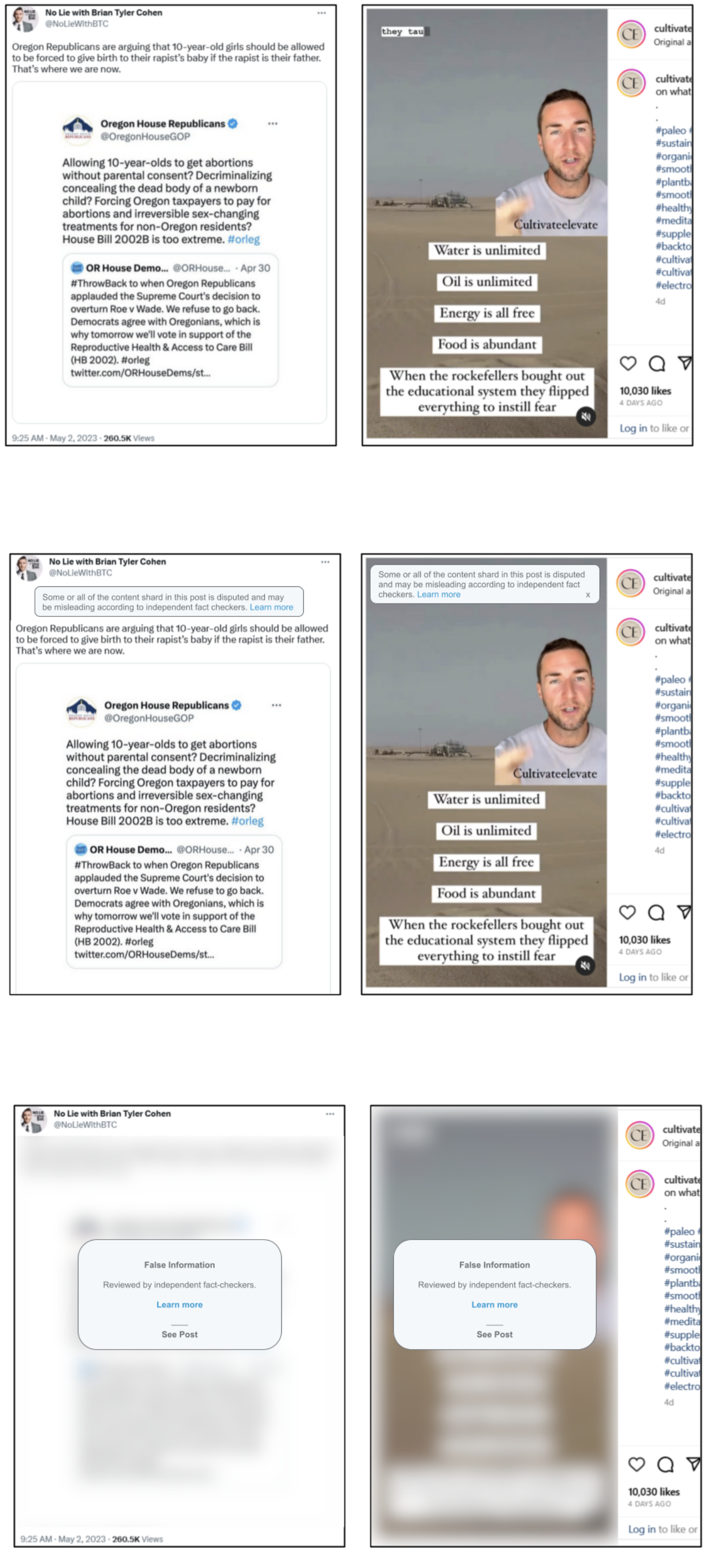

Figure 1: The left column shows text posts and the right column shows video posts. The top row is control group, middle row is banner flag, and bottom row is blurred screen with flag.

In all conditions, the participants will be shown 5 posts in total. Two of them are misinformation posts and three are filler posts (featuring non-malicious content like animals). The order of those five posts will be randomized for each participant, and the video and text content counterparts will be created to be as similar as possible.

Lastly, I carefully selected two reference misinformation posts that I felt would be representative of misinformation more broadly. I factored in picking one post that comes from a conservative view and another that comes from a liberal speaker to keep our choices balanced in terms of political affiliation. Next, the two posts cover two topics: climate and abortion, which are both political topics, and that helps this experiment generalize to hot topic conversations. These posts are also good options because they have been fact checked by a reputable source (factcheck.org), and they pertain to recent news events. Here is a quick summary of each reference post:

Abortion: A liberal social media post mischaracterizes conservative opposition to an Oregon bill by claiming Republicans said girls "should be allowed to be forced to give birth to their rapist's baby if the rapist is their father" [7].

Climate denialism: a TikTok video shared on Instagram falsely claims that there is an "unlimited" supply of oil, and people are being "taught" otherwise to keep them "in a fear state" [8].

4.2 Outcome Variables and Regression Analysis

Replicating the methods of Taber and Lodge (2006) [9], I will measure my outcome variable, belief in misinformation, using survey questions that address attitude position and attitude strength. **Attitude position** will be measured using a binary scale with the question "To the best of your knowledge, is the claim in the previous post about (abortion/climate) accurate?" And the answer options: yes or no. **Attitude strength** will be measured on a 100 point-sliding response scale with the question "How confident are you in your answer?" And answer option what is a sliding scale from 0 to 100, where 0 = Not confident at all, 100 = extremely confident. I choose this method over a 5-point Likert scale because it allows us to capture a finer grain analysis of people's beliefs.

Then, I would run three regressions in the form

$$Y = \beta_1 D_1 + \beta_2 D_2 + \beta_3 (D_1 \times D_2) + \gamma X + \epsilon$$

where D_1 represents a categorical variable for whether the content shown is text or video based, D_2 is a categorical variable for whether the participant was shown control, banner flag, or blur with flag, and X represents the other covariates included to help control for confounding variables such as demographic data.

(Epsilon represents irreducible error.) Using this format, $\beta_1, \beta_2, \beta_3$ represents the treatment effects I am interested in.

Where the three regressions differ is in the Y variable. In one of the regressions, Y would be a value capturing the attitude position averaged across the the two misinformation posts, where 1 = yes, belief in misinformation and -1 = no, no belief in misinformation. In the second regression, Y would be the attitude strength, which is a variable lying in the [1, 100] interval. Lastly, in the third regression, Y would equal the product of the two outcome variables described in the previous sentences to capture the interaction between the attitude position and attitude strength.

Since this is a standard randomized control trial, I can simply estimate standard errors using the standard output from running "lm" in R.

4.3 Additional Survey Design Points

Attention Checks: I will add three screener questions that put a subtle instruction in the middle of a block of text to check for attention (model after Lee and Bissell [5]).

Pre-survey: I would also want to collect and screen general demographic data and their political affiliation and prior beliefs about abortion and climate change before the treatment (which would be incorporated in the X variable in the description of the regressions above).

5 Validity

5.1 Threats to Internal Validity and Potential Fixes

Failure of randomization: To make sure that the treatment is well randomized, I will use a balancedness table to check what the split looks like and fix anything that looks off before implementing the treatments.

Attrition: Because my hypothesis is that video posts are more engaging than text posts, a plausible concern is that a higher proportion of participants in the text group will end up not completing the survey than the participants in the video group. I could potentially fix that by splitting 60% people into text and 40% into video, instead of doing a 50/50 split.

5.2 Threats to External Validity and Potential Fixes

Non-representative sample: One big concern is that with an online survey, people self-select into participating in the survey and being in a survey requires the participant to be 1) tech savvy enough to find them, 2) care enough (even with incentives like pay given to the participants), and 3) have the free time to do a random survey. People with these attributes are not necessarily representative of the entire adult population in the US so this is a threat to the external validity of this experiment.

Non-representative program: Another potential issue is that my treatment only involves scrolling through 5 social media style posts, whereas in reality, many people scroll past well over 5 posts when they are on social media. This discrepancy means that my experimental treatment is not necessarily representative of the real-world treatment. The potential fix for this point and the one above is to simply run additional experiments with wider coverage in the future.

Hawthorne effect: People have been shown to behave differently if they know that they are being studied. It is possible that participants will be thinking more critically about whether or not they believe what they saw during the experiment than they usually do when encountering information in the real world because they know it is a study. As such, the synthetic simulation may fail to represent participants' authentic online behavior, which may be better captured in a naturalistic setting. This effect should not directly confound or compromise the predicted outcomes of the experiment, but it is something to keep in mind when interpreting the results.

6 Conclusion

This study will help improve public understanding of the downstream consequences of video misinformation, considering their differences compared to text-based misinformation. As a result, the findings will contribute to creating an actionable framework to design effective platform nudges and flags to identify video disinformation and strategies to detect and counter the spread of video-based misinformation. While there are still limitations that this proposed study will not be able to address such as concerns regarding the freedom of speech being obstructing and resistance from platforms, this work would still be a very important contribution to this developing field of literature.

References

[1] S Shyam Sundar, Maria D Molina, and Eugene Cho. Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication*, 26(6):301–319, 2021.

[2] Rui Hou, Veronica Perez-Rosas, Stacy Loeb, and Rada Mihalcea. Towards automatic detection of misinformation in online medical videos. *2019 International Conference on Multimodal Interaction*, 2019.

[3] Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, and Ryota Kobayashi. Modeling the spread of fake news on twitter. *PLOS ONE*, 16(4), 2021.

[4] Paul Mena. Cleaning up social media: The effect of warning labels on likelihood of sharing false news on facebook. *Policy & Internet*, 12(2):165–183, 2020.

[5] Jiyoung Lee and Kim Bissell. User agency–based versus machine agency–based misinformation interventions: The effects of commenting and ai fact-checking labeling on attitudes toward the covid-19 vaccination. *New Media & Society*, page 14614448231163228, Apr 2023.

[6] Chloe Wittenberg, Ben M. Tappin, Adam J. Berinsky, and David G. Rand. The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences*, 118(47):e2114388118, Nov 2021.

[7] Nora Macaluso. Viral video makes false claim about global oil supply, May 2023.

[8] Saranac Hale Spencer. Post mischaracterizes gop opposition to oregon bill on reproductive health care for minors, May 2023.

[9] Charles S. Taber and Milton Lodge. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769, 2006.