Graduate Capstone Report

Kendall Beaver

2025-05-13

Intro

This report contains the assessment and critical analysis of 594 written responses that were collected from 99 participants via a questionnaire in Dr. Diana Daly's pilot study, "Design of audio ads to prebunk misinformation and promote civil discourse", where participants were asked the following qualitative questions ("QQ"):

- QQ1: What are some reasons you think people are attracted to radical groups?
- QQ2: What are some reasons you think people believe false ideas circulating online?
- QQ3: Describe a situation in which supporting friends would be more important than telling the truth.

These questions were asked before and after an ad was presented to a participant, which could be one of four ads: a control ad in an audio or visual-audio format, or a test ad in an audio or visual-audio format. The former ad was a random ad, whereas the latter was an experimental ad designed to increase participants' information literacy and make them more skeptical about misinformation. In addition to QQ1 - QQ3, participants were asked what they liked and disliked about their ad, plus any additional thoughts they had about said ad.

Thematic and content analysis was performed on the responses to identify common theme. Every response was thoroughly reviewed several times, and if the meaning of a response still could not be understood after this process, the response was placed in the "Other" category—different from the "Unsure" category, where participants explicitly stated that they were unsure and did not have an answer for a QQ.

A composite score was also developed to measure how much a person's response changed to each QQ, by taking the count of words and sentences before and after an ad, normalizing the counts to make them equivalent, adding them together, then dividing by two to produce an average score.

$$\frac{\text{Normalized count of words} + \text{Normalized count of sentences}}{2}$$

In summary, participants' views remained mostly unchanged before and after an ad, except for QQ1, which had 9 categories (the most amount of categories for any QQ) and the largest amount of change. Additionally, the composite score always increased whenever participants were show an audio-visual ad, whether a control or test ad, and when the data was analyzed based on gender instead of ad group, there were slight contrasting views in how males and females answered a question.

This report is organized and presented in four main sections: I. Categorical Responses for QQ1 - QQ3; II. Composite Score for QQ1 - QQ3; III. Sentiment Analysis of Ads; and IV. Future Work & Acknowledgment.

I. Categorical Responses for QQ1 - QQ3

QQ1

A total of 9 common themes were identified in the responses for QQ1, the largest amount of themes for all QQs, and this could be due to the nature of the question—trying to understand the thoughts of somebody different than the participant and why they would be interested in joining a radical group. If a theme appeared at least 3 times then it was assigned its own category, but many themes were combined together if they had an influence on one another, helping to produce a more concise list.

"What are some reasons you think people are attracted to radical groups?"

1. Coerced into joining group
2. Influenced to join a group in a non-coercive manner
3. Drawn to narcissists; to people who are passionate, confident, and charismatic
4. Confirmation bias; wanting to be part of a community of like-minded people, to not feel lonely
5. Feeling helpless and believe in the group's cause; that they can make change and life better
6. Curious about checking out the group; group seems large, powerful and different; could be fun, could gain status; to not feel bored
7. Are narcissists themselves and want to influence or have power over people; enjoy expressing their views
8. Unsure why people are attracted to radical groups
9. Other; response is too broad, unclear, and/or doesn't match any other categories

Figures 1 & 2 below show the results of these categories changed after presenting an ad

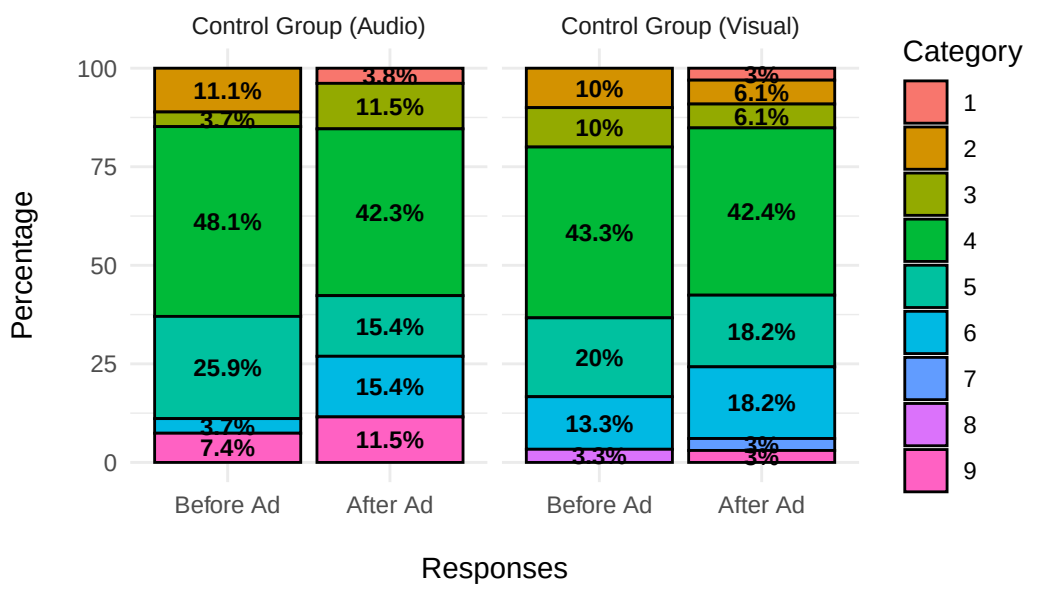

Figure 1: Control group (visual & audio) responses to QQ1 before and after ad.

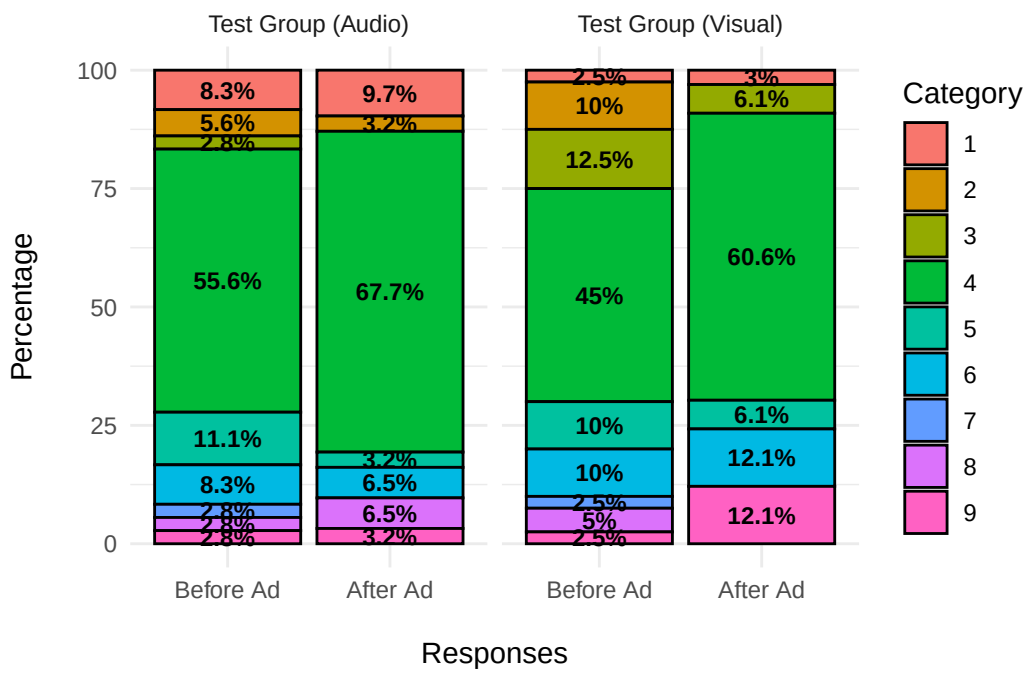

Figure 2: Test group (visual & audio) responses to QQ1 before and after ad.

The largest category for both ad groups was Category 4 ("Confirmation bias"), which composed over 40% of all responses for participants. However, this response decreased on average by 7.07% after the control ad was presented, whereas responses increased by 28.21% on average after the test ad was presented to those participants.

In the latter ad, a female character named Mel finds a group of online friends who want to storm the post office, but in the former ad a female character named Jessica doesn't say that her friends are radical in any way, only that they're "early bowlers." It's possible that the ad with radical friends could have influenced the responses, but follow-up survey data and/or interviews need to be conducted to validate this.

The second largest category for the control ad was Category 5 ("Believe group can make their life better") before the ad, at 25.9% for the audio group and 20% for the visual group, then was tied and shared equally with Category 6 ("Group seems large, powerful and different; could be fun") post-ad, at 15.4% for the audio group and 18.2% for the visual group (Figure 1).

Regarding the audio ad, Category 5 was the second largest category for the audio group, at 11.1%, while Category 3 ("Drawn to narcissists") held the same ranking for the visual group, at 10%. But after the ad was presented, Category 1 ("Coerced into joining group") took second place for the audio group, at 9.7%, while Categories 6 & 9 ("Other; response is too broad") equally shared second place for the visual group, at 12.1% (Figure 2).

QQ2

6 main themes were identified for QQ2 and appeared in all groups except for Category 4 ("Unsure"), which appeared once for the test ad - visual group, as shown in Figure 4.

"What are some reasons you think people believe false ideas circulating online?"

1. Doesn't take time to fact check

2. Believes first thing they see
3. Looks for ideas that confirm pre-existing beliefs
4. Unsure
5. Other
6. There's too much misinformation available

Categories 1 ("Doesn't take time to fact check"), 2 ("Believes first thing they see") & 3 ("Looks for ideas that confirm existing beliefs") occupied the first three places across all ad groups, but in different variations. As shown in Figure 3, Category 3 held the top spot at 32.1% for the control ad - audio group, but was replaced by Category 1 post-ad, at 33.3%. For the visual group, Categories 1 & 2 shared first place at 31%, until Category 1 became the top response post-ad, at 38.7%.

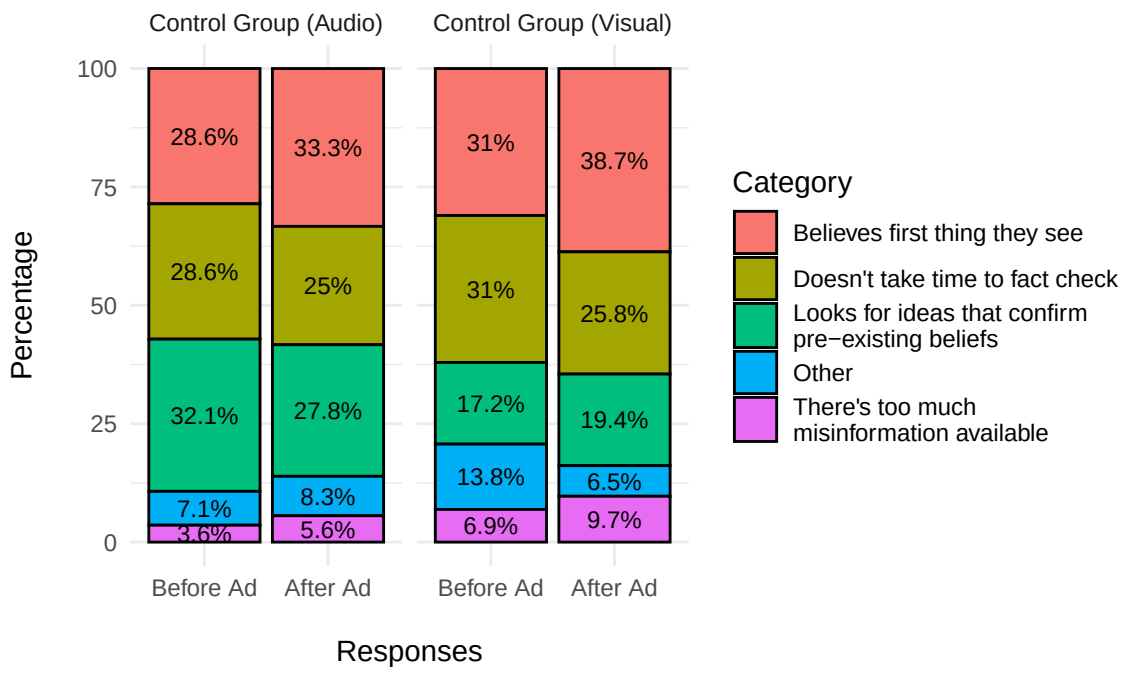

Figure 3: Control group (visual & audio) responses to QQ2 before and after ad.

For the test ad, Category 1 was the dominant category for the audio and visual group, at 34.4% and 43.2%, respectively, and remained in the top spot for the visual group post-ad, increasing to 47.5%, but decreased to 26.3% for the audio group post-ad and was overtaken by Category 2, which gained a slight lead of 28.9% (Figure 4).

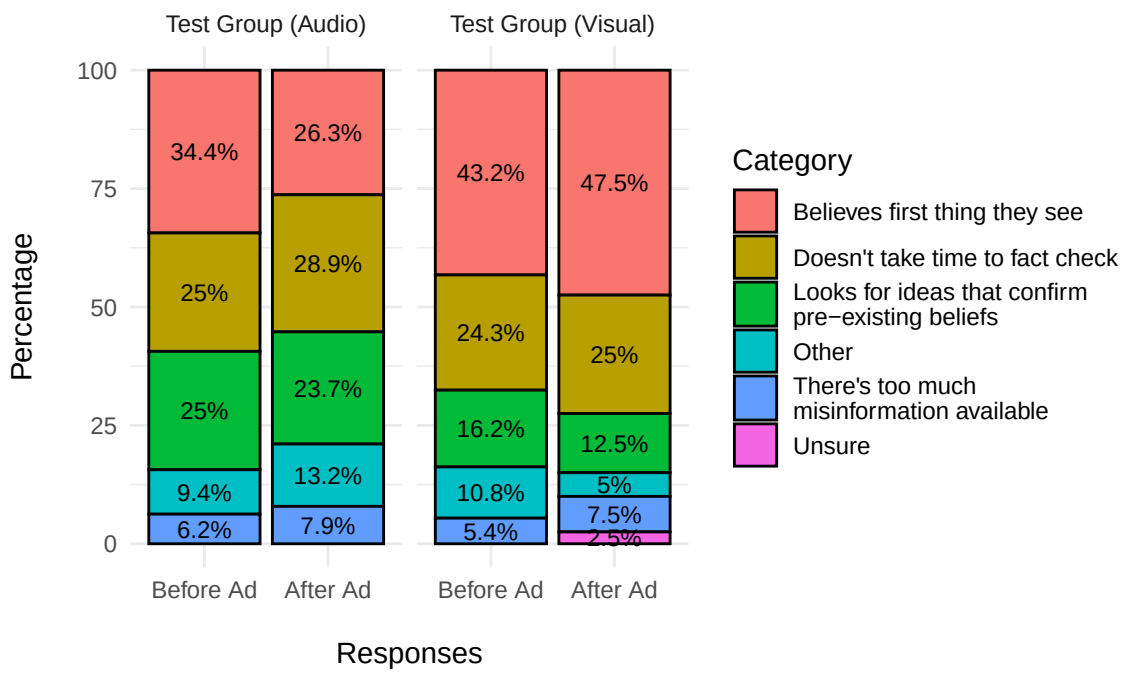

Figure 4: Test group (visual & audio) responses to QQ2 before and after ad.

QQ3

This question had the least amount of categories for all QQs, yet contained the most interesting discovery when the data was split by gender instead of ad group. Before the gender findings are discussed, the 4 main categories identified in QQ3 will be discussed and presented first.

"Describe a situation in which supporting friends would be more important than telling the truth."

1. Protect friend from danger and/or punishment
2. Unsure
3. White lie
4. Would tell truth over supporting friends

Unlike the previous QQs which asked open-ended questions, QQ3 asked participants to describe a scenario in which lying would be okay to support friends. All of the responses could fit into 1 of 4 categories, indicating that there isn't a variation with lying in this type of situation, and none of the answers were ambiguous enough to create an "Other" category, making this the only QQ to not have that category.

Figure 5 shows that the largest response for all groups before and after the ad, overwhelmingly, was Category 3 ("White lie"), hovering between 62% - 82% before the ad and 67% - 82% after the ad. Category 1 ("Protect friend from danger and/or punishment") remained the second largest category for 3 out of 4 groups (control ad - audio & visual groups; test ad - visual group), ranging from 13.6% - 21.4%, while Category 4 ("Would tell truth over supporting friends") took second place for the test ad - audio group at 14.3% before the ad, and sharing this same percentage and ranking with Category 1 post-ad.

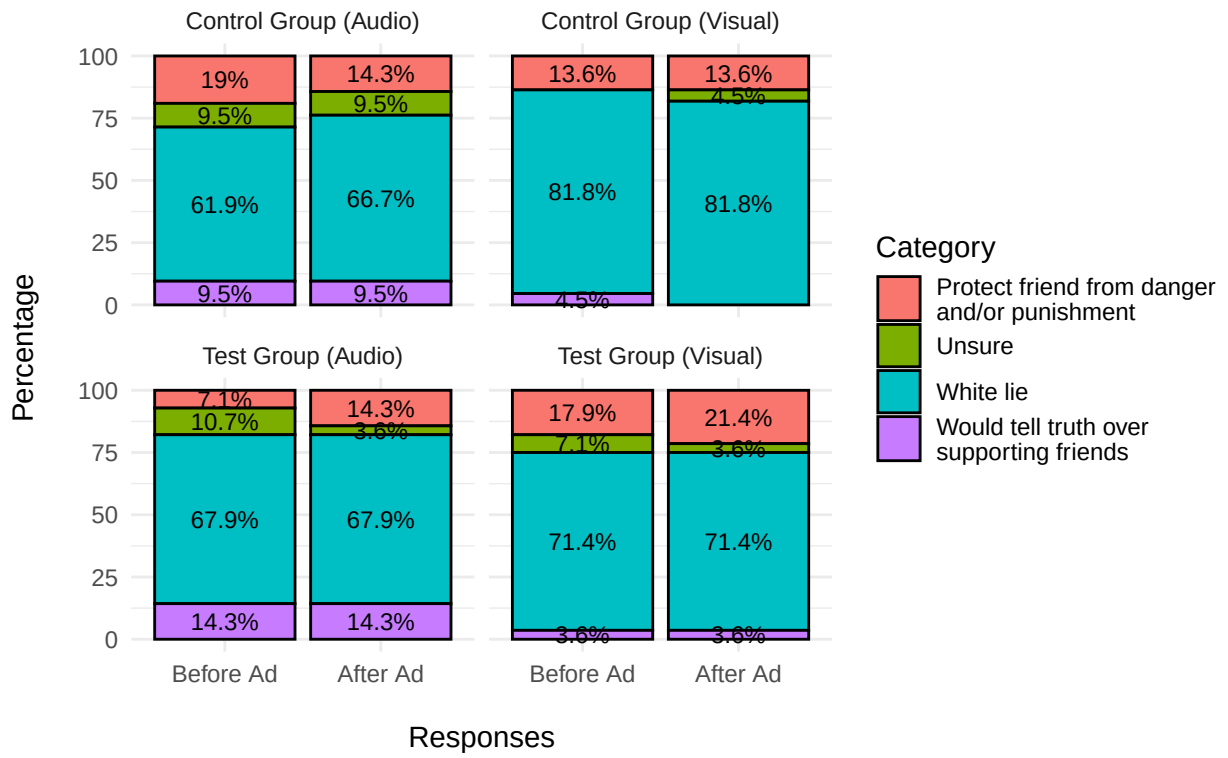

Figure 5: All group responses to QQ3 before and after ad.

Moving onto the gender findings, it must first be noted that only females and males were included in this analysis because only 2% of participants identified with the third category, "prefer not to say", whose responses heavily skewed and shifted the data before and after an ad. Considering these two genders, females are approximately 3.25 times more likely than men to tell the truth over supporting their friends, whereas males are approximately 44% more likely than females to protect their friends (Figure 6). Overall, each gender kept the same responses for QQ3 pre- and post-ad, and agreed to nearly the same percentage, around 70%, that it's okay to tell a white lie to protect a friend.

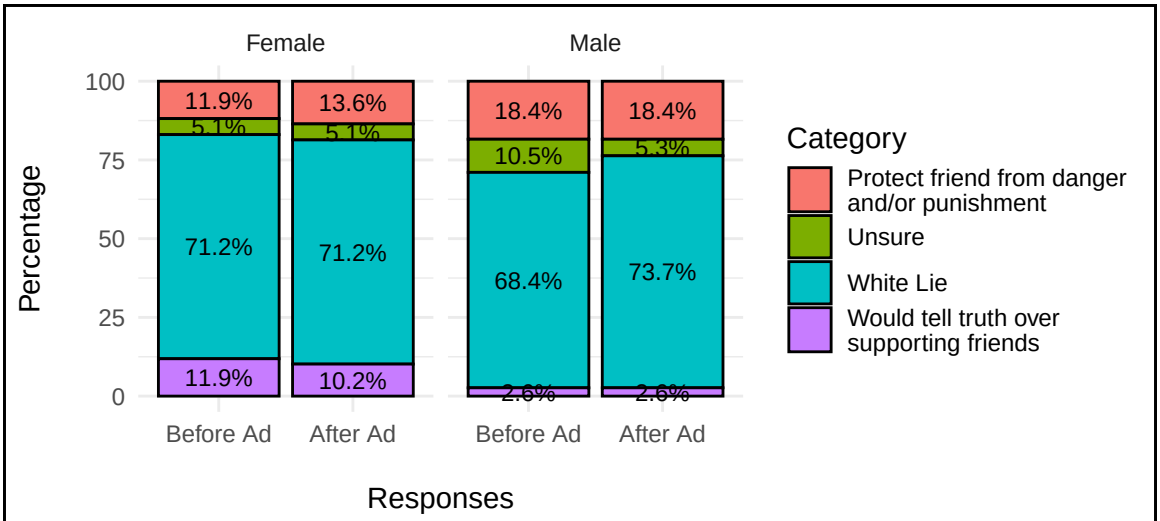

Figure 6: Female and male responses to QQ3 before and after ad.

II. Composite Score for QQ1 - QQ3

As mentioned in the Introduction, a composite score was developed to objectively measure how much a participant's response increased or decreased after being presented with an ad. This score was first used to quantify how much a response remained the same, in which an arbitrary range from -25% to 25% was chosen to encapsulate a participant keeping the same number of thoughts with some flexibility, because it's impractical to expect someone to memorize and use the exact number of words and sentences as before the ad.

These results are presented in Table 1, where the ad that appears to have little effect on changing the responses was the control ad - audio group for QQ2, where 76.2% of responses remained unchanged. In contrast, the most amount of responses that had the least amount of change was the for same ad, but for the visual group in QQ3, where 31.8% of the responses remained unchanged.

Table 1: Responses post-ad that remained the same, between -25% to 25%.

GROUP	QQ1	QQ2	QQ3
Control Group (Audio)	52.4	76.2	52.3
Control Group (Visual)	50.0	54.5	31.8
Test Group (Audio)	57.2	42.9	50.0
Test Group (Visual)	53.6	46.4	39.3

To see the percentage of responses that decreased over 50%, Table 2 was created. The test ad - audio group had the highest percentages for both QQ1 and QQ3, with 21.4% and 10.7%, respectively. Additionally, the same ad was correlated with the visual group for having the highest percentage for QQ2, with 14.3%.

Table 2: Responses post-ad with a decrease greater than -50%.

GROUP	QQ1: < -50%	QQ2: < -50%	QQ3: < -50%
Control Group (Audio)	14.3	9.5	9.5
Test Group (Audio)	21.4	10.7	10.7
Test Group (Visual)	10.7	14.3	3.6
Control Group (Visual)	NA	9.1	4.5

Lastly, to see which score had a significant impact on double the responses to at least over 100%, Table 3 was created for such an analysis. The control ad - test group showed the greatest proportion for QQ1, at 27.3%, while the test ad - visual group had the highest proportion for QQ2, at 14.3%, and both the test ad - audio and visuals group tied for the highest proportion of responses in QQ3, at 7.1%.

Table 3: Responses post-ad that increase by 100% or more.

GROUP	QQ1: >= 100%	QQ2: >= 100%	QQ3: >= 100%
Control Group (Visual)	27.3	13.6	4.5
Test Group (Visual)	10.7	14.3	7.1
Test Group (Audio)	NA	3.6	7.1

The results from these tables show many inconsistences, where certain formats performed well with a certain group but then performed poorly with the same group for the same QQ. A good example of this is for QQ2 (Tables 2 & 3) where the test ad - visual group had the highest proportion of responses (14.3%) for increasing responses at least over 100%, but it also had the greatest decrease in response length by at least over 50% (14.3%). However, there are many factors that could have influence how long a participant made their response post-ad, such as a long delay between answering the QQs or using GenAI to help form their answers, and therefore, additional data is needed to make more conclusive inferences.

III. Sentiment Analysis of Ads

The following analysis was conducted by analyzing the responses from the Ad Response Questions (ARQs) in the survey:

- ARQ 1: What did you like most about the ad you just watched?
- ARQ 2: What did you like least about the ad you just watched?
- ARQ 3: Which words or phrases, if any, in the ad that you just watched made you feel particularly good or bad?
- ARQ 4: Please tell us any other thoughts you have about the ad you just watched.

ARQs 1 & 2 was easy to determine the sentiment that a participant felt because the question poses a participant to feel a certain sentiment, while ARQs 3 & 4 could express any sentiment; these latter two required the most consideration and analysis to determine and fully understand the sentiment they were feeling. A new category called "SENTIMENT" was created that categorized the overall sentiment that was expressed most frequently between ARQs 1-4, and was identified as 1 of 3 categories: Positive, Neutral, or Negative.

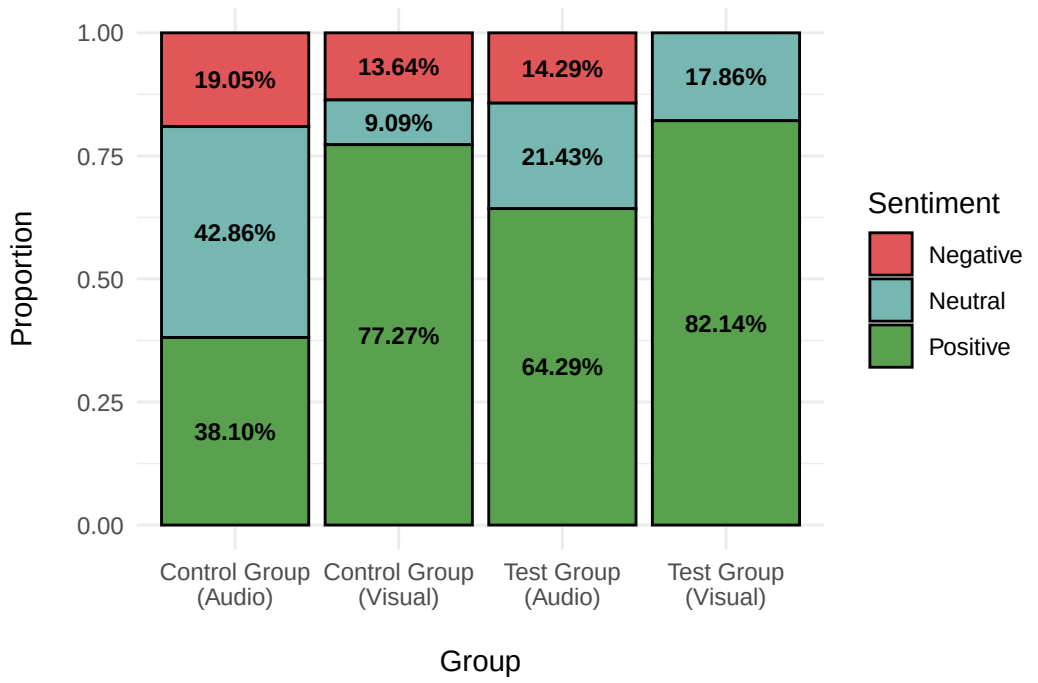

Figure 7: Sentiment towards ad based on group.

The overall sentiment categories have been plotted in Figure 7, revealing that the control ad - audio group performed the worse out of all groups, 19.05%, followed by the test ad - audio group at 14.29%. This dislike for audio ads aligned with the behavior discovered from podcast listeners: Less than half of all participants, 47.5%, reported listening to at least one podcast, and out of those listeners, 71.8% always skip the ads compared to 1.4% who claimed to never skip ads.

In contrast, the test ad - visual group reported the highest positive sentiment, 82.14%, followed by the control ad - visual group, 77.27%, where the former group was the only group to report no negative sentiment towards the test ad in its visual-audio format.

It's possible that the participants in the audio format groups may have had a preconceived negative view towards audio ads if podcast listeners skip them at a very high rate, while participants could have a preconceived positive view towards visuals ad due to "Shorts" videos that appear on social media platforms, the same format as the visual ads where words of dialogue appear across the screen, and therefore feels more familiar. But a follow-up survey and/or interview would need to be conducted to verify such claims.

IV. Future Work & Acknowledgment

The categories that were created for QQs 1-3 can be used as a template to form a drop-down list of responses that a participant can choose from, selecting the category that best aligns with their thoughts instead of writing out a response, saving time for the participant and allowing the researcher to perform a quick, quantitative analysis on the data. However, the limitation of using these categories as-is is that they were determined by one person, although done so with great consideration and critical thought, but at least one more annotator should assess and form categories of their own, then compare both lists to form a more objective list.

Additionally, it is recommended that QQ1 be reduced from 9 categories to around 5 categories so as not to overwhelm a participant, which aligns with the number of categories for QQs 2 & 3, 6 and 4 categories, respectively. But even in this case, the ad may never influence a response from pre- to post-ad, and so the categorical responses may best serve as survey preliminary questions only.

One thing not included in this report is a section regarding the evaluation of GPT-2 to help with Dr. Daly's recent study, "From libraries to 'truth factories': Epistemic performance in prebunking sketches with AI chatbots," as discussed in the project proposal. I recommended using GPT-2 because it is open-source and has many parameters that can be adjusted by a user, which I thought could adjust the level of humor and produce a more comedic, GenAI-created scripts for advertisements, compared to GTP-4 which is not open-source and incapable of receiving any input for how responses are produced. However, after running many tests, GPT-2 performed horrendously, creating scripts that were composed of random words and formed no logic in the narrative, even after adjusting and testing different values for the parameters. This analysis therefore was quickly discarded and not included in this final report.

The original R markdown file and code for this report, plus relevant data and files are located *https://github.com/kendall-beaver/misinformation-study,* which the people involved in this study only have access to.

I would like to express my gratitude to Dr. Diana Daly and her undergraduate research assistant Kainan Jarrett for allowing me to be a part of their project, post-study, which was more interesting and fun than I ever expected. Because of their research and guidance, I not only increased my information literacy and skepticism for misinformation, but I've become a better qualitative researcher for life. I am more than grateful for this unique, eye-opening, and insightful learning exp