Do Female YouTubers Receive More Negative Comments Than Male YouTubers?
A Literature Review and Computational Study

Comp400 - Final Paper

Grace Yuzwin, 261150398

August 4, 2025

Abstract: The stagnation and decline in gender equality today (Equal Measures 2030, 2024) is concerning. With YouTube being the second most popular social network with over 2 billion monthly users as of February 2025 (Dixon, 2025), the comments received on YouTube can help indicate if the popular view on gender differences reaches into social media. This study consolidates the existing literature and further investigates if female YouTubers receive more negative comments than male YouTubers, in order to see if gender inequality is being perpetuated on the YouTube platform. Using selected pairs of one female and one male in the same YouTube category with similar channels, along with the YouTube Data API v3 and the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool in two custom Python scripts, all of the comments from the selected channels most recent 10 videos were analyzed. The methodology used concludes that gender does not have a substantial impact on the number of negative comments received, but that male YouTubers receive more negative comments than female YouTubers by a small (0.22%) margin. Given the limitations outlined in this study, the conclusion should be interpreted with caution and is not definitive.

Table of Contents

Introduction

"NO COUNTRY IS ON TRACK TO ACHIEVE GENDER EQUALITY BY 2030" (Equal Measures 2030, 2024), and if current trends continue, a girl born today will have to live to the age of 97 to experience a society without gender inequality. With YouTube being the second most popular social network with over 2 billion monthly users as of February 2025 (Dixon, 2025), having over one billion hours of videos being watched per day (Goodrow, 2020), and over 20 million videos uploaded on the platform every day (YouTube, n.d.), the user interaction on YouTube through the comments is a tool to monitor the public's opinions, and how they compare to the current state of gender equality.

In this study I attempt to consolidate and focus the previous studies on the differing negative comments received by female and male YouTubers by creating a more defined selection and comparison criteria, as well as utilizing the YouTube Data API v3 and the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool to allow for more accurate analysis.

Literature Review

A growing body of research has examined gender-based disparities in how content creators are treated on social media platforms. YouTube, as a highly visible and comment-driven space, has become a site for investigating how male and female content creators receive different types of comments. This literature review examines six major studies on this topic, highlighting their findings, methods, and relevance to the current research, independent of each other.

One of the first works to examine gender differences in the comments received by YouTubers was Wotanis and McMillan (2014). They argued that Jenna Mourey (Jenna Marbles) received more negative comments on her videos than her male counterpart, Ryan Higa. Wotanis and McMillan manually analyzed the most recent 100 comments on the top ten most popular videos from each YouTuber's channel. They found that Mourey received a higher proportion of critical and hostile comments (18%) compared to Higa (4%), as well as more inappropriate sexual comments. The argument from Wotanis and McMillan align with the focus of the present study, but the methodology differs, as their data was collected and analyzed manually rather than through automated sentiment analysis. While they were able to provide valuable early insight into gendered comment disparity, their manual collection causes their analysis to be limited by its small sample size.

Then, in 2018, Döring and Mohseni replicated the previous study done by Wotanis and McMillan to gain a deeper insight into male dominance and sexism online, while generalizing their findings. Firstly, they analysed the 100 most recent comments from the 12 most popular videos by Mourey and Higa, finding that Mourey received roughly three times more negative feedback than Higa. Secondly, 5 pairs of comparable female and male comedy YouTubers from North America were selected and their 100 most recent comments from their 6 most recent videos were manually extracted and analysed. They found that women receive less negative feedback (11%) than men (14%), but that women receive less positive (52%) feedback than men (54%). While their intention to recreate and generalize Wotanis and McMillan's study aligns with the focus of the present study, the manual extraction and analysis methodology, as well as their choices of comparable pairs of female-male YouTubers differ.

Döring and Mohseni continued in 2020 by doing another systematic replication of the original study done by Wotanis and McMilan in 2014 to see if their findings extend across other YouTube genres and video platforms. They selected the four most popular YouTube genres in Germany and one female and one male German-speaking YouTuber in the same genre. They then extracted the 100 most recent comments from the 10 most popular videos from each YouTuber manually and analysed them. This led to findings that confirmed the findings of Wotanis and McMilan in 2014, while contradicting their own previous 2018 study, and concluding that women have a higher risk of becoming the victims of gendered online hate speech on YouTube. But, they found a low prevalence of online hate speech in general since hate comments can be eliminated by moderators and users on YouTube. Similarly to their previous study, their arguments align with the focus of the present study, but the manual extraction and analysis methodology differ.

With a focus on science education, Saiz et al. (2023) explored Spanish-speaking, active, and large (100,000 or more subscribers) science YouTube channels to identify whether there are differences that could deter women from becoming professional popularisers of science. They used the web scraping software Octoparse to collect and export a maximum of 100 random comments per video per YouTuber and manually evaluated each comment for sentiment. They found that "…popular science YouTube channels hosted by women produce, in relative terms, a greater number of both positive and negative personal comments." While the purpose of the study and manual analysis does not align with the focus of the present study, the use of a web scraping software aligns.

Similar to the purpose of Saiz et al. (2023), Veletsianos et al. (2018) explored how the gender of a video presenter (YouTuber) influences the sentiment of comments and subsequent replies, to gain a better understanding of the range of sentiments scholars may face online. Web extraction and data mining methods were used, along with the YouTube API, to gather the comments to analyse, then SentiStrength was used to determine the sentiment of the comments and replies. They found that videos of male YouTubers showed greater neutrality, while videos of female YouTubers saw greater positive and negative polarity. The methodology and argument used closely resemble the methodology and argument used in the present study.

Most recently, Weismann et al. (2025) investigated how socio-structural characteristics, including gender, shape differences in the occurrences of hate speech on YouTube. While in pre-print, Weismann et al. still provide a broad and expansive exploration of these effects, looking into 3,695 channels, 430,000 videos, and roughly 40,000,000 comments total. They compiled the comments using the YouTube Data API v3 and the comments were then analyzed using the XGBoost algorithm and a multilingual BERT model, both trained on 7,500 manually annotated German and English comments. They found that female YouTubers receive significantly more positive sentiment and are exposed to significantly fewer publicly visible offensive comments compared to male YouTubers, but, with a noted limitation on the impact of YouTube's tools against problematic content and hate speech. The methodology used is in line with the present study, and while the argument is broader, it is similar.

Taken together, the studies reveal a pattern of disproportionate negativity in comments directed at female YouTubers, particularly in the form of hostility and sexualization, with YouTube moderation being a common limitation. However, the research varies in its methodology, ranging from small-scale manual analysis to machine learning-driven sentiment classification. While some studies relied on subjective interpretation, others introduced algorithmic models trained on large datasets. Notably, few studies directly compare creators of similar popularity or content style, leaving room for further investigation. The present study builds on this literature by applying a consistent algorithmic approach to multiple matched pairs of YouTubers, aiming to provide a more controlled comparison of gendered comment dynamics.

Methodology

To achieve the goal of detecting if female YouTubers receive more negative comments than male YouTubers, firstly, YouTubers were selected in pairs of one female and one male in the same YouTube category and with similar channels to attempt to isolate the comment differential by gender. Secondly, to collect and analyse enough comments, all comments on the 10 most recent videos (as of July 31st) from each YouTuber were retrieved and analysed, attempting to improve the existing methodologies, by having the capacity to not limit the number of comments analysed due to the VADER sentiment analysis tool and the YouTube Data API v3.

YouTuber Selection

As YouTube doesn't publicly release precise, category-specific video counts, choosing categories based on popularity was not an option. Instead, the YouTube website "Explore" sidebar subcategory was used to evaluate for current recommended categories, which were the same across all accounts. At the time of evaluation (last confirmed on July 31st, 2025), the sidebar contained: Music, Movies & TV, Live, Gaming, News, Sports, Learning, Fashion & Beauty, Podcasts, and Playables. Music, Movies & TV, Live, and Playables were removed from evaluation due to not having any comments on the content or having limited commented videos, then News and Learning were removed due to the heavy effects the specific content of the videos have on the comments. The category of Comedy was added due to its involvement in a majority of the literature previously discussed.

For each category, a pairing of one female and one male YouTuber was chosen, and to help ensure that any negative comments directed at these two YouTubers cannot be attributed to differences in content or category, but rather stem from gender-based bias, the channels had to be "similar". This means that they have either similar subscriber counts or similar view counts on their 10 most recent videos and they share audiences, with either being known collaborators or having similarly titled content.

To ensure proper gender confirmation, channels were manually verified for being centered around the gender associated with the primary channel holder. If the channel was run by an individual, the gender of only the individual was confirmed by either using their YouTube channel description or other social media platforms. If the channel was operated by a group, all members of the group must be of the same gender, and that was confirmed either using their YouTube channel description or other social media platforms.

10 channels were selected from the 5 different categories, see Appendix B for specific channels.

Collection and Evaluation of Comments

In order to properly collect and evaluate the comments, custom Python scripts were written to automate the process. The code is divided into two main files:

- "comment-to-spreadsheet.py", which takes a YouTube playlist ID of 10 public videos (in this case the 10 most recent videos per channel) and with the available YouTube Data API v3 (Google Developers, n.d.), the public comments from all 10 videos are extracted along with their like counts, and saved to a CSV (Comma Separated Values) file. All comments from the 10 most recent videos from the selected YouTubers were extracted on July 31st, 2025.
- "spreadsheet-to-VADER.py", which takes the CSV file containing YouTube comments and their like counts, created by "comment-to-spreadsheet.py", and uses the VADER sentiment analysis tool to evaluate the sentiment of each comment, then scales the sentiment based on the like count. It returned both scaled information on the compound VADER score averages and a CSV file containing the comments and their VADER compound score.

Full source code and comment analysis CSV files are available on GitHub (see Appendix A for the repository link).

The VADER sentiment analysis tool was selected to evaluate the comments because it is open-source, easy to integrate, and specifically designed for analyzing social media text, as well as performing well on short, informal texts (Hutto & Gilbert, 2014), making it suitable for YouTube comments. Additionally, the compound scoring system offers a straightforward way to quantify overall sentiment in a single value, aligning well with the goal of this project.

As implemented in "spreadsheet-to-VADER.py", each comment's sentiment was evaluated using the compound VADER score, "normalized to between -1 (most extreme negative) and +1 (most extreme positive)... which is a normalized, weighted composite score" (Hutto, 2022). As for the sentiments of a compound score, a positive sentiment is evaluated to have a compound score of >= 0.05, a negative sentiment is evaluated to have a compound score of <= -0.05 and a neutral sentiment is evaluated to have a compound score between (not inclusive) 0.05 and -0.05, as explained by Hutto (2022).

Sentiment scores were scaled by like count, treating each like as an endorsement of the comment's content. In this approach, every like a comment receives is treated as if the comment were submitted again, effectively multiplying the sentiment impact of that comment. This method assumes that liking a comment signals agreement with its sentiment and message, and thus serves to reinforce and amplify the original statement's influence in the overall sentiment analysis.

Results

As seen in Figure 1, gender does not have a substantial impact on the amount of negative comments received. It shows that female YouTubers receive more negative comments in the Comedy and Gaming categories, but male YouTubers receive more negative comments in the Podcast, Sports, and Fashion & Beauty categories. Figure 1 also shows that the average sentiment of the comments is lower for female YouTubers in the Comedy, Gaming, and Sports categories, but male YouTubers receive lower average sentiment scores in the Podcast and Fashion & Beauty categories.

Category	Name	Gender	# Positive Comments	# Negative Comments	# Neutral Comments	Total Comments	Average Sentiment	% Negative Comments	Expected % Negative
Comedy	Quenlin Blackwell	Female	370925	226031	162692	759648	0.157	29.75	4
Comedy	LARRAY	Male	569080	226737	188421	984238	0.2095	23.04	3
Podcast	BroskiReport	Female	146027	42906	54341	243274	0.2635	17.64	6
Podcast	Very Real Good	Male	60449	35158	29464	125071	0.1344	28.11	2
Gaming	GeminiTay	Female	164669	92296	51899	308864	0.1644	29.88	6
Gaming	EthosLab	Male	212633	68842	75357	356832	0.275	19.29	5
Sports	Celine Dept	Female	51255	2337	20364	73956	0.435	3.16	14
Sports	Dude Perfect	Male	240144	26252	42969	309365	0.4741	8.49	2
Fashion & Beauty	Lydia Tomlinson	Female	4197	282	161	4640	0.6918	6.08	4
Fashion & Beauty	Tim Dessaint	Male	12683	1597	4100	18380	0.4027	8.69	6

Figure 1: VADER sentiment analysis summaries of the collected comments, along with the expected percentages of negative comments, based on manual evaluation of the top 100 comments from each YouTuber's CSV file (see Appendix A for details).

Combined with Figure 2, which shows that the female and male YouTubers only have a 0.22% difference in the number of negative comments received, with female YouTubers receiving fewer negative comments, and there is only a 0.043 difference in the average compound sentiment of the comments received by male and female YouTubers, again female YouTubers receiving a higher average sentiment, it can be concluded that gender does not have a substantial impact on the amount of negative comments received.

Gender	Negative Comment %	Average Sentiment
Female	17.30231729	0.34234
Male	17.52287061	0.29914

Figure 2: Condensed VADER sentiment analysis grouping together the analysis by gender.

Limitations

With the current state of YouTube moderation, automatic detection systems are constantly scanning comments, and if they are found to violate the YouTube Community Guidelines, the comment is automatically removed. There are also options for regular users of YouTube to report comments and options for channels to hold potentially inappropriate comments for review (Google, 2025). These options for automatic and manual removal of negative comments significantly impact the results shown above, as any comment containing hate speech, harassment, or cyberbullying is automatically removed by YouTube, and any comment deemed inappropriate can be deleted by the channel owner. The removal of these comments directly impacts the results found and may even invalidate them, as the present study does not question YouTube's moderation, but the public's interaction with content produced by different genders.

In addition, during the process of comment sentiment analysis, many mismatches between sentiment score and intended meaning were observed, causing incorrect sentiment averages and negative comment percentages, which greatly impact the results. This can be seen by comparing the "% Negative Comments" and the "Expected % Negative" columns in Figure 1, noting that the large difference in values can be attributed to VADER's incorrect analysis.

Below are three examples with explanations as to why the compound sentiment score is incorrect:

1. "Larray and vanilla are deadly together. 😭😭😭😭😭"

This comment received a compound score of -0.9382, and while it can be understood that the word deadly and crying emojis have a negative connotation, in this case, this sentence implies that the commenter enjoys the content and thinks the two mentioned are funny, actually praising the two. This comment should have received a positive compound score.

2. "How can people hate on quen😭😭😭😭"

This comment received a compound score of -0.9442, and while it contains the word hate and crying emojis, this is again a situation where, on their own, the words can have negative connotations, but together in a sentence, this commenter is actually enjoying the content and cannot see how someone else could not.

3. "An april's fools update episode would go crazy!"

This comment received a compound score of -0.7088, and while there is improper grammar, that is not a reason for a negative compound score. This comment is supportive of the content and thinks that an update episode for April Fools' Day is a good idea, which should have given the comment a positive compound score.

Conclusion

The methodology used concludes that gender does not have a substantial impact on the amount of negative comments received, but that male YouTubers receive more negative comments than female YouTubers by a small (0.22%) margin. Given the limitations outlined, the conclusion is not definitive and cannot be used to monitor the public's opinions, and how they compare to the current state of gender equality.

Future Works

Access to removed and deleted comments is necessary, and a more accurate sentiment analysis tool should be used in comment evaluation. It is also important to expand into and include other genders in evaluation, not just female and male.

Furthermore, it would be interesting to examine how real-world gender views are reflected, or whether they are reflected at all, in YouTube comments. A numerical analysis of the impact of moderation on negative, gendered comments on the platform would also be valuable.

References

Cambronero-Saiz, B., Cristófol-Rodríguez, C., & Segarra-Saavedra, J. (2023). Content analysis from a gender perspective of comments received by Spanish science youtubers. *Media and Communication*, *11*(1), 252–263. https://doi.org/10.17645/mac.v11i1.5971.

Dixon, S. J. (2025, March 26). *Biggest social media platforms by users 2025*. Statista. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ .

Döring, N., & Mohseni, M. R. (2020). Gendered hate speech in YouTube and Younow comments: Results of two content analyses. *Studies in Communication and Media*, *9*(1), 62–88. https://doi.org/10.5771/2192-4007-2020-1-62.

Döring, N., & Mohseni, M. R. (2018). Male dominance and sexism on YouTube: results of three content analyses. *Feminist Media Studies*, *19*(4), 512–524. https://doi.org/10.1080/14680777.2018.1467945.

Goodrow, C. (2020, July 18). *You know what's cool? A billion hours*. YouTube Official Blog. https://blog.youtube/news-and-events/you-know-whats-cool-billion-hours/ .

Google Developers. (n.d.). *YouTube Data API v3*. Google. https://developers.google.com/youtube/v3.

Google. (2025). *YouTube's Community Guidelines*. Google. https://support.google.com/youtube/answer/13209064?hl=en&ref_topic=3014329&sjid=10887659745551958741-NA.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550.

Hutto, C. (2022, April 1). *VADER-Sentiment-Analysis: README*. GitHub. https://github.com/cjhutto/vaderSentiment.

Equal Measures 2030. *2024 SDG gender index*. Equal Measures 2030 . (2024, December 9). https://equalmeasures2030.org/2024-sdg-gender-index/.

Veletsianos, G., Kimmons, R., Larsen, R., Dousay, T., & Lowenthal, P. (2018). Public comment sentiment on educational videos: Understanding the effects of presenter gender, Video format, threading, and moderation on YouTube TED talk comments. *PLOS One*, *13*(6). https://doi.org/10.1371/journal.pone.0197331.

Weismann, S., Philipp, A., Verwiebe, A., Krauter, C. O., Fritsch, N.-S. & Buder, C. (2025). *Clicks, comments, consequences: Are content creators' socio-structural and platform characteristics shaping the exposure to negative sentiment, offensive language, and hate speech on YouTube?* arXiv. https://doi.org/10.48550/arXiv.2504.07676.

Wotanis, L., & McMillan, L. (2014). Performing Gender on YouTube: How Jenna Marbles negotiates a hostile online environment. *Feminist Media Studies*, *14*(6), 912–928. https://doi.org/10.1080/14680777.2014.882373.

YouTube. YouTube by the Numbers. YouTube for Press. (n.d.). https://blog.youtube/press/

Appendix A: Supplementary Materials

GitHub Repository: https://github.com/graceyuz/COMP400

This repository includes:

- Python scripts for data collection and sentiment analysis.
- CSV files of YouTube comments, their like counts, and their VADER compound scores organized per YouTuber.
- README documentation.

Appendix B: YouTuber Pairings

Category	Female YouTube Channel	Male YouTube Channel
Comedy	Quenlin Blackwell	LARRAY
Podcast	BroskiReport	Very Really Good
Gaming	GeminiTay	EthosLab
Sports	Celine Dept	Dude Perfect
Fashion & Beauty	Lydia Tomlinson	Tim Dessaint

Figure 3: YouTuber female-male pairings by category.