5-2022

Pognalysis: An Analysis of Gender and Language on Twitch.tv

Julia Michelle Cardillo

Pognalysis: An Analysis of Gender and Language on Twitch.tv

Julia Michelle Cardillo

Dr. Kier Hanratty

Pace University, Department of Economics

Presentation Date: March 26th, 2022

Graduation Date: May 16th, 2022

Abstract

With the success of the video game and eSports industries, comes the rise of Twitch.tv (also known as Twitch), the most popular live-streaming platform today. Being that the most popular activity to stream on Twitch.tv is video games, there is a lot of overlap between those who play video games or are avid eSports fans and those who watch Twitch. However, these communities are notorious for their hostility towards women. The goal of this study is to assess, if any, the differences in atmosphere, as well as the differences in the presence of gendered language/harassment, profanity, and non-gaming/non-activity related words between male and female Twitch streamers. In this study, I web-scraped live Twitch chats, pulled statistically overrepresented words, and ran various probit regressions to accomplish this. Findings indicate that there is somewhat a presence of gendered language/harassment and profanity on Twitch.tv. There is also a slight but statistically significant difference in probability of profanity, and environment between male and female Twitch streamers. The specific marginal effects of gender showed that for a male streamer, the likelihood of presence of harassment or profanity increased by 1.26%, while the likelihood of presence of harassment decreased by .08% in comparison to female streamers. Lastly, for a male streamer, the likelihood of presence of profanity increased by 1.42% in comparison to female streamers. Harassment, however, showed to be statistically insignificant, possibly due to various caveats.

Table of Contents

List of Figures and Tables

Introduction

Founded in 2005 as a start up called Justin.tv, Twitch.tv (also known as Twitch) is the largest live-streaming platform on the internet, with 41.5 million viewers as of 2020 (eMarketer Editors, 2020). On Twitch.tv, users can stream practically anything, the most popular activity being video games. There is also a live chat in which viewers can interact with or donate to the stream. Today, Twitch has an abundance of categories to browse, including everything from your favorite video game to even "Pools, Hot Tubs, and Beaches"!

Twitch's success is partially attributed to the rapidly growing video game and eSports industries—there are 227 million gamers in America alone as of 2021 (Entertainment Software Association, 2021). Additionally, the global eSports audience in 2021 reached approximately 474 million (Takahashi, 2021). This considered, with the continued success of Twitch.tv and the rise of the video game and eSports industries, more visibility has been brought to one of the most persistent issues in video game culture: sexism. Sexism is not at all new to the communities of the video game and eSports industries, as women have typically been harassed and alienated from competitive eSports, casual play, and even the workplace (Browning, 2021; Gardner, 2021; Kuznekoff & Rose, 2012; Limbong, 2022).

So why does this matter? Being a woman who has spent her whole life playing video games, and has experienced this harassment and alienation first hand, I am interested in assessing current moderation practices, as well as understanding language used within these communities. In this paper, I seek to investigate the difference in atmosphere, language, profanity, and harassment for male and female streamers on Twitch.tv. The main questions I seek to answer are:

- Is the atmosphere created by female Twitch streamers distinct from male Twitch streamers?
- Is there more non-activity/non-gaming related or gendered language/harassment in streams by female Twitch streamers than male Twitch streamers?

In this study, through use of statistically overrepresented words, detection of harassment and profanity from lexicons, and probit model regressions, I discovered that there is gendered language/harassment on Twitch. Additionally, I discovered that there is a slight difference in chat environments fostered by male and female Twitch streamers, in which the amount of profanity and harassment differs slightly.

I believe that researching these questions is necessary in order to create a more inclusive environment on these platforms for *all*. In this paper, I will first analyze previous literature relating to my topic. Next, I will hold a pre-estimation discussion and then discuss the details of my data and methodology. I will then go over the results of my research and make conclusions based on my findings.

Literature Review

As mentioned previously, instances of harassment and alienation are not new to the video game and eSports industries. Being that these industries go hand in hand, it is no secret that these issues exist on Twitch as well.

Male-Dominated Industry

The video game and eSports industries, on both sides, prove to be male dominated industries. This is evident in creation and development, as well as audience. In fact, a 2021 game developer survey found that 61% of game developers were male, while 30% were female (International Game Developers Association, 2021). This gender disparity creates negative externalities in the environments in which these games are developed, creating an extremely hostile, and even deadly work environment for women.

A recent example of this is the Activision Blizzard lawsuit, in which Activision Blizzard paid an $18 Million settlement to employees who were subject to “servere sexual harassment and pregnancy discrimination” (Limbong, 2022). Employees who spoke up against this matter were retaliated against and supressed. In the lawsuit, one example was given of a female employee who “died by suicide during a business trip, as a result of her sexual relationship with her male supervisor. Before her death, male colleagues allegedly shared explicit photos of the woman” (Browning, 2021). The Activison Blizzard lawsuit is just one of many examples of how dangerous these environments can be for women to exist in, and they are still coming forward with their stories of harassment and minimization today (Gardner, 2021; Lorenz & Browning, 2020). Another infamous example of this is Gamergate.

Gamergate began in August of 2014, as a series of “rape and death threats, the hacking of personal accounts, and the release of personal information (doxing)” towards game developer Zoe Quinn (Buyukozturk, Gaulden, & Dowd-Arrow, 2018). These violent and misogynistic attacks occured in response to claims made by Quinn’s ex-boyfriend, that her game *Depression Quest* had only received good reviews due to her intimate relationships with game journalists (Massanari, 2016). This response to the "perceived lack of ethics within gaming journalism" then progressed to target a wide range of women and events promoting diversity in the industry (Buyukozturk et al., 2018). Gamergate is just one of the many instances that reveal that in the video game and eSports industries, women's accomplishments are usually invalidated, and that hostility and violence towards them is a normal occurrence. The prognosis of the future, however, seems to be brighter, being that the number of female game developers has increased by 6% since 2019, and that every day more light is being shed on the injustices of the industries (International Game Developers Association, 2021). However, despite this increase, it is clear that this is still a male dominated field all-around, even in audience. For example, a 2021 study conducted by the Entertainment Software Association (ESA) found that 45% of gamers identify as female, while 55% identify as male[1] (Entertainment Software Association, 2021). This all considered, the disproportionate environments seem to influence the actual content of the games, as well as the environments in which these games are developed and played.

Objectification and Sexualization of Women

Due in part to the disproportionate amount of female game developers to male game developers, the content of video games themselves often underrepresent or portray harmful stereotypes of women (Gestos et al., 2018; Kondrat, 2015). In video games, women are often portrayed as objects of sexual fetishization, subordinate to the male characters, and are rarely the main character (Dietz, 1998; Fox et al., 2014); this is even true for video game advertisements as well (Behm-Morawitz, 2014; Scharrer, 2004). In most cases, if a

[1] It is important to note however, that this study gathered data from about four thousand Americans. So while it is normally distributed, it might not be fully representative, seeing as it doesn't show the percentage of non-binary or gender non-conforming gamers. For the purpose of my study, however, I am mainly concerned regarding male-to-female ratios.

female character is in a video game (especially as the main character), chances are she has very little clothing and large breasts, like *Tomb Raider*'s Lara Croft. Lastly, research from various studies shows that these representations may feed into negative self-images for women, and have an association to real-life perceptions and hostility towards them (Dietz, 1998; Fox et al., 2014; Fox & Potocki, 2015; Gestos et al., 2018). All of these factors may contribute to the real-life harassment women experience in these communities.

Gendered Language and Harassment

Various literatures discuss the trials of being a woman in the video game and eSports communities, as well as the use of gendered language on Twitch. For example, a study by Ruvalcaba et al. (2018) found that female streamers received more sexual harassment comments than male streamers, pertaining to the streams of the most watched competitive eSports titles. In addition, a study performed by Nakandala et al. found that streamer gender is associated with the kind of messages they receive. In this study, they also pulled the most statistically overrepresented words from the stream, which inspired me to recreate this for my study (Nakandala et al., 2016). Another study performed by Farrell et al. found that harassment and violence towards women online have increased (Farrell et al., 2019). In this study, they utilized different lexicons to detect the presence of misogynistic rhetoric on Reddit (Farrell et al., 2019). This inspired me to use words from these lexicons for my study. In Kuznekoff and Rose, it was revealed that in an online multiplayer game setting, a female voice prompted three times as much negative feedback than a male voice (Kuznekoff & Rose, 2012). Lastly, women who "achieve a moderate level of competence are rendered invisible or are actively marginalized" (Paaßen et al., 2016). In other words, due to the overall misogynistic environment of these communities, as well as the overrepresentation of men as the highly visible and professional figures in gaming culture, women are often forced to create their own communities within gaming, hide their identity, or step away from gaming (Ruvalcaba et al., 2018). This online hostility and harassment occurs often and is something I even have experienced many times while playing games online.

Pre-Estimation Discussion

After analyzing previous literature, I believe that in the presence of a female streamer, more harassment or profanity will be present. In addition, I predict that there will also be more non-activity/non-gaming related words present in the chat, being that women in the video game and eSports communities are often objectified or marginalized. I believe this will happen because of the portrayal and underrepresentation of women in video games, as well as the effect of this on women's self-image and others' perception of women.

Data and Methodology

Question:

- Is the atmosphere created by female Twitch streamers distinct from male Twitch streamers?
- Is there more non-activity/non-gaming related or gendered language/harassment in streams by female Twitch streamers than male Twitch streamers?

Hypothesis:

I hypothesize that in the presence of a female streamer, there will be more harassment, profanity, or non-activity related comments (referring to the activity in the stream) in the chat.

Variables:

- **Male -** Indicates whether or not the streamer is male or non-male (1 and 0 respectively). For the scope of this study, non-male just means female. This variable is my variable of interest (or main independent variable), to which I want to test if there is any causal relationship between this and the probability of gendered language/harassment comments.
- **Number of Followers -** The streamer's number of followers in millions.
- **Genre_e -** The activity happening or game being played during the stream. Includes: Apex Legends, Among Us, Chess, GTA V (Grand Theft Auto V), Just Chatting, League of Legends, and VALORANT.
- **Time of Stream -** Day, Night, Overnight. Day (7am-5pm), Night (5:01pm-12am), Overnight (12:01am-6:59am).
- **Harassment -** Presence of harassment in observation (or comment) (1 or 0).[2]
- **Profanity -** Presence of profanity in observation (1 or 0).
- **Harpro -** Presence of harassment in comment (1 or 0).

Table 1: Summary Statistics

	mean	sd	min	max
harpro	.0122883	.1101709	0	1
profanity	.0120426	.1090772	0	1
harassment	.0006881	.0262238	0	1
genre_e	4.293716	1.664827	1	7

[2] Words used for Harassment, Profanity, and Harpro were selected from Hatebase, and Offensive/Profane Lexicon.

dt_e	1.750768	.7183063	1	3
channel_e	8.246012	3.727981	1	15
followers	3.600683	3.045029	.141	8.4
N	40689			

Table 1 shows the summary statistics for all of my variables. Due to the mainly categorical nature of my data, most of the summary statistics are not useful. However, it is an interesting feat to note that the mean amount of followers among all of the streamers studied was 3.6 million. In addition, it is important to note that my sample size was 40,689 observations (or comments), making my data normally distributed.

Methodology:

For the data generating process, I generated the data on my own. The data I collected was from *live* Twitch streams. There were 7 different games/activities (I will refer to these just as activities) from which I collected: Among Us, Apex Legends, League of Legends, VALORANT, Grand Theft Auto V, Just Chatting, and Chess. For each of these activities, I pulled 30 minute intervals from one male and one female streaming that respective activity. This was so that I could ensure that I was covering a broad range of categories, as well as compare the differences in chat between the different genders. In order to find which streamer I would pull from, I clicked on the respective category on Twitch and located whichever streamer came up that had at least over 100 thousand followers. This was to ensure that I would have normally distributed results, as each chat comment counts as its own observation, thus eliminating any sample size issue.

From there, I web-scraped the chat log from these live channels. To do this, I used a script (https://www.learndatasci.com/tutorials/how-stream-text-data-twitch-sockets-python/) made by Brendan Martin, with his permission. I tweaked the code slightly, as I was running into some errors with printing. In addition, I also altered the code to put the pulled chat into a Pandas dataframe, making it easily transferable to Microsoft Excel. These scripts were written in Python, and in order to run this, I used the environment Pycharm. This script pulls the exact date and time of the comment, which channel it was pulled from, the username of the commenter, and the content of the comment. The dates and times at which I pulled these streams were at random, but since I tested the marginal effect and significance of this in my regression, I was not concerned with pulling out a consistent date and time.

I spent well over 6 hours collecting data, as it all had to be manually collected, and there was a good chunk of data I did not end up using. After my data was collected, I then compiled all male data and all female data into different excel files, to be processed with Python Pandas. To process this data, I created a script that pulls the 200 most common words present in the dataframe. From there, I manually removed prepositional words, as well as streamer specific words because the external validity of this is faulty. In addition, the script detects the presence of Profanity or Harassment in each observation, returning a 1 or 0 next to the observation. It is important to note that I also created a variable called Harpro. Harpro is the union of Profanity and Harassment, and returns whether the observation contains harassment OR profanity words. After collecting this data, I then looked directly at the streamer's Twitch page to get the number of followers and added this to my Excel file. In addition, I added the streamer's gender as M (Male) or F (Female). One main caveat to my data collection was that my computer buffer size was too small, so I could not collect more than 30 minutes worth of stream time per streamer. This also means that my sample size is not as large as I wanted it to be.

Below shows the probit models I ran. In addition, I also ran marginal effects for these probit models.

- **$harpro_i = \beta 1male_i + \beta 2genre\_e_i + \beta 3dt\_e_i + \beta 4followers_i + \beta 5channel\_e_i + u_i$**
- **$profanity_i = \beta 1male_i + \beta 2genre\_e_i + \beta 3dt\_e_i + \beta 4followers_i + \beta 5channel\_e_i + u_i$**
- **$harassment_i = \beta 1male_i + \beta 2genre\_e_i + \beta 3dt\_e_i + \beta 4followers_i + \beta 5channel\_e_i + u_i$**

I chose to run various probit models, due to the nature of my data being mainly categorical, and especially my left hand side variables being binary. By using a probit model, I was able to test for the change in probability of Harassment or Profanity being present. All of the independent variables are the same throughout each model and include the variables male, genre_e, dt_e, followers, and channel_e. I made genre_e and dt_e into indicator variables so that I could assess the marginal effects of each activity, as well as each time of day.

Results and Discussion

Below are the results from my various probit regressions:

Table 2: Probit Regression Results	(1)	(2)	(3)
	Harpro	Harassment	Profanity
male	0.362^{***}	-0.281^{*}	0.416^{***}
	(0.0644)	(0.169)	(0.0678)
1.Apex Legends	0	0	0
	(.)	(.)	(.)
2.Among Us	0	0	0
	(.)	(.)	(.)

3.Chess	-1.246***	0	-1.203***
	(0.227)	(.)	(0.228)
4.GTA V	-0.475***	-0.283	-0.515***
	(0.114)	(0.305)	(0.121)
5.Just Chatting	-0.214***	3.060	-0.192***
	(0.0625)	(132.3)	(0.0644)
6.League of Legends	-0.263***	3.381	-0.270***
	(0.0880)	(132.3)	(0.0887)
7.VALORANT	-0.445***	0	-0.467***
	(0.119)	(.)	(0.124)
1.Day	0	0	0
	(.)	(.)	(.)
2.Night	-0.133**	0.0102	-0.130**
	(0.0583)	(0.206)	(0.0586)
3.Overnight	-0.0622	3.343	-0.0636
	(0.0997)	(132.3)	(0.103)
channel_e	0.00416	-0.0330	0.0105
	(0.0114)	(0.0213)	(0.0123)
followers	-0.0488***	-0.0618*	-0.0488***
	(0.0168)	(0.0369)	(0.0173)
_cons	-1.999***	-5.615	-2.103***
	(0.122)	(132.3)	(0.131)
N	35372	30250	35372

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Pseudo R^2:
Harpro: 0.0392
Harassment: 0.0452
Profanity: 0.0429

My probit regressions returned slightly unexpected results, and my pseudo R^2s are definitely not as high as I had hoped. I also cannot directly interpret the coefficients from the probit model. However, it is important to note the signs of the coefficients, as the positive coefficient is not what I expected for my male (or gender) variable in relation to Harpro. This means that in the presence of a male streamer, the likelihood of presence of harassment or profanity in the chat increases. For my various categories or games, the coefficients show to be negative in comparison to my reference variable which is Apex Legends. For the time of day, comparative to my reference variable Day, the coefficients of Night and Overnight show to be negative. Additionally, the followers variable shows to have a negative coefficient. Now, you may notice that the coefficients for Among Us and Apex Legends show up as zero; I will explain this after the marginal effects.

Below is my marginal effects table:

Table 3: Marginal Effects Table	(1)	(2)	(3)
	harpro	harassment	profanity
male	0.0126^{***}	-0.000855	0.0142^{***}
	(0.00229)	(0.000534)	(0.00237)
1.Apex Legends	0	0	0
	(.)	(.)	(.)
2.Among Us	0	0	0
	(.)	(.)	(.)
3.Chess	-0.0267^{***}	0	-0.0259^{***}
	(0.00370)	(.)	(0.00372)
4.GTA V	-0.0189^{***}	0	-0.0193^{***}
	(0.00458)	(.)	(0.00455)
5.Just Chatting	-0.0108^{***}	0	-0.00967^{***}
	(0.00363)	(.)	(0.00364)
6.League of Legends	-0.0127^{***}	0	-0.0126^{***}
	(0.00425)	(.)	(0.00417)
7.VALORANT	-0.0182^{***}	0	-0.0182^{***}
	(0.00474)	(.)	(0.00472)

1.Day	0	0	0
	(.)	(.)	(.)
2.Night	-0.00465**	0.0000203	-0.00446**
	(0.00210)	(0.000410)	(0.00207)
3.Overnight	-0.00234	0.397	-0.00234
	(0.00366)	(36.67)	(0.00367)
channel_e	0.000145	-0.000100	0.000359
	(0.000398)	(0.0000668)	(0.000419)
followers	-0.00170***	-0.000188	-0.00167***
	(0.000588)	(0.000117)	(0.000596)
N	35372	30250	35372

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For organizational purposes, this section will be broken down by the Harpro, Harassment, and Profanity probit models respectively. On another note, Apex Legends, and Among Us, and Day show up as zeros across the board. For Apex Legends, this is a reference variable for the genre_e variable. This means that every coefficient for genre will be interpreted in comparison to Apex Legends. So, the coefficient for Chess is interpreted as such: compared to Apex Legends, Chess sees a 2.67% decrease in likelihood of harassment or profanity. For Among Us, however, it is not a reference variable. So why does it also show zeros across the board? The answer to this, in my opinion, makes a lot of sense. Because Among Us shows up as zeros, and considering it is not the reference variable, this means that the Python script did not pick up any instances of Harassment, Profanity, or Harpro. This might be due to the fact that Among Us is a multiplayer strategy game that is loved by people of all ages, especially children. It would also make sense that this lack of detection is due in part to strong moderation and filtering in the chat—I will explain more about this later. Lastly, Day shows up as zeros because it is a reference variable for the dt_e variable. I have also decided to skip the analysis of channel_e, being that it is not a main focus of my study, and that I did not make it an indicator variable when running regressions.

Harpro

Male: As shown in the table, in the presence of a male streamer, the likelihood of Harpro (harassment or profanity) increases by 1.26%.

Chess: Compared to Apex Legends, Chess sees a 2.59% decrease in likelihood of harassment or profanity.

GTA V: Compared to Apex Legends, GTA V sees a 1.93% decrease in likelihood of harassment or profanity.

Just Chatting: Compared to Apex Legends, Just Chatting sees a 1.08% decrease in likelihood of harassment or profanity.

League of Legends: Compared to Apex Legends, League of Legends sees a 1.27% decrease in likelihood of harassment or profanity.

VALORANT: Compared to Apex Legends, VALORANT sees a 1.82% decrease in likelihood of harassment or profanity.

Night: Compared to Day, Night sees a .5% decrease in likelihood of harassment or profanity.

Overnight: Compared to Day, Overnight sees a .2% decrease in likelihood of harassment or profanity.

Followers: With every one follower increase in the number of followers, there is a .1% decrease in the likelihood of harassment or profanity.

The negative coefficients for all of the activities/genres make sense to me, being that Apex Legends is a First Person Shooter, and by nature is a much more violent game. The rest of the categories, besides VALORANT and GTA V are a lot less violent in nature. The coefficient for the male variable was expected.

Harassment

Male: As shown in the table, In the presence of a male streamer, the likelihood of harassment decreases by .08%

Night: Compared to Day, Night sees a .002% increase in likelihood of harassment.

Overnight: Compared to Day, Overnight sees a 39.7% increase in likelihood of harassment or profanity.

Followers: With every 1 follower increase in the number of followers, there is a .01% decrease in the likelihood of harassment or profanity.

For harassment, the sign of the coefficient was as I predicted, however, I believe that there were various caveats affecting the results of my data. For instance, due to moderation practices on Twitch, as well as streamer filtered words, the Harassment words were not detected in various genre_e activities. This also might be due to the fact that the words chosen from the lexicons may not be the best words to use for the study, as Twitch has very specific slang and language that may not be used on other sites. Lastly, these coefficients for harassment showed to be statistically insignificant, which I wasn't expecting.

Profanity

Male: As shown in the table, In the presence of a male streamer, the likelihood of profanity increases by 1.42%.

Chess: Compared to Apex Legends, Chess sees a 2.59% decrease in likelihood of profanity.

GTA V: Compared to Apex Legends, GTA V sees a 1.93% decrease in likelihood of profanity.

Just Chatting: Compared to Apex Legends, Just Chatting sees a .96% decrease in likelihood of profanity.

League of Legends: Compared to Apex Legends, League of Legends sees a 1.26% decrease in likelihood of profanity.

VALORANT: Compared to Apex Legends, VALORANT sees a 1.82% decrease in likelihood of profanity.

Night: Compared to Day, Night sees a .4% decrease in likelihood of profanity.

Overnight: Compared to Day, Overnight sees a .2% decrease in likelihood of profanity.

Followers: With every one follower increase in the number of followers, there is a .1% decrease in likelihood of profanity.

As mentioned earlier, the zeros for Among Us and Apex Legends make sense due to the aforementioned reasons. However, I was somewhat surprised to see that the presence of a male streamer increases the likelihood of profanity being present. Lastly, the coefficients for the different genre_e activities do not surprise me, as by how violent Apex Legends is in nature.

My results, in general, differ slightly from my expectations, but this study is definitely not without caveats. I wasn't expecting to see so many zeros for harassment, but I believe this is due to one of the main caveats of my study: current Twitch moderation practices.

Currently, Twitch has various moderation practices in place for use by streamers. There is a page on Twitch.tv called Creator Camp that lists the various ways streamers can moderate their chat. One of the main methods is AutoMod, which "uses machine learning and natural language processing algorithms to hold potentially inappropriate or offensive messages from the chat. The creator or a channel moderator can review them before appearing to other viewers in the chat." (Twitch.tv, n.d.). In addition to this, streamers usually have dedicated moderators, who review and scan the chat as comments come in and approve or deny comments held by AutoMod. I believe this greatly impacted the results of my data being that there are many moderation practices in place, so many of the words I am looking for may be caught in AutoMod, and won't even reach the chat.

Statistically Overrepresented Words / Qualitative Study

Below are tables for the ten most statistically overrepresented words in female and male streams, respectively.

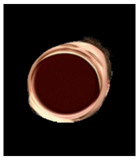
Tables 4 & 5: Statistically Overrepresented Words

Female - Words	Occurrences
invite	853
pog	410
omegalul	328
kekw	316
lul	279
hair	272
<3	236
pepelaugh	228
lulw	202
lol	188

Male - Words	Occurrences
kekw	1808
lul	935
omegalul	522
letsgo	486
lol	470
lmao	316
ayaya	279
pepela	279
sadge	279
pogu	273

These results are not surprising to me, as they both contain a large amount of the same commonly used Twitch emotes, with a few exceptions. The most popular emotes pulled from the observations are “pog”, “omegalul” and “kekw”.

For context and clarity, here are some of the statistically overrepresented emotes and their definitions:

- Pog: A derivative of PogChamp, used to convey excitement.
- Kekw: Used to represent laughter.
- Omegalul: Used to represent laughter.
-  Lul: Indicates laughter.

Generally speaking, these emotes all convey some type of laughter, with different undertones; all are very common on Twitch. This shows that the environments are somewhat similar but still differ nonetheless.

However, it is very interesting to note that one of the most statistically overrepresented words for female streamers was “hair”. This may be due to the fact that I pulled from one of Pokimane’s streams. Pokimane is an extremely popular female streamer and is usually seen with straightened hair. She is of Moroccan descent, and as of recently, she decided to start streaming with her natural, curly hair texture. After finding all of the observations that contain “hair”, most of them show to be overwhelmingly positive. However, as expected there are various hate comments that say “ugly hair”, and “your hair is better straight”. Additionally, some of the comments regarding hair seem to be microaggressive (even if unintentional), referring to her natural hair as her “real” hair.

“<3” also appears as one of the most statistically overrepresented words, which is a text variation of displaying a heart. After looking at the data, this shows to be a way to convey love in a cutesy way. Additionally, by looking at the observations a lot of the comments containing “<3” were part of a chatbot message that congratulates viewers for upkeeping their subscriptions to the streamers. The use of this is overall very positive.

Conclusion

To summarize, in this study I utilized statistically overrepresented words, detection of harassment and profanity from lexicons, and probit model regressions, to analyze gendered language/harassment, profanity, non-gaming/non-activity related words, and chat environments on Twitch. More specifically, I wanted to address, if any, the distinctions of this between male and female Twitch streamers.

According to my results, there is somewhat a presence of gendered language/harassment and profanity on Twitch.tv. Additionally, there is a slight but statistically significant difference in probability of profanity and environment between male and female Twitch streamers. Harassment, however, showed to be statistically insignificant.

Although I ran into issues detecting specifically harassment words in my data set, **I firmly believe that this harassment still exists online—but that we might just have efficient methods of chat moderation.** As I mentioned before, my data is not without caveats, and each streamer's individualized moderation practices most definitely could have affected the outcome of my data. Furthermore, I believe that most of these comments exist, but they are either caught before they even reach the chat (via chat filtering) or that moderators remove them promptly as they come in. This would be especially true for the more popular streamers, who might want to be cognizant of a younger audience. Lastly, it could have also been the type of words I pulled from the lexicons, considering that Twitch has a very specific language and slang that might not be used on other sites.

There are many things that I would like to do in future research. This involves studying the effects of time of day on harassment and profanity. While I did somewhat analyze the results of this variable, it was not my main focus, so the dispersion of stream times is uneven. In addition, I would love to collect a larger sample size if possible, as well as alter some of the lexicon words chosen in my script to minimize my caveats. Lastly, I would like to look more into non-activity or non-gaming related words, as my study did not provide much to work with. Ultimately, I hope that my research (and furthering my research) will provide valuable insight that will aid in striving for a safe and inclusive environment online, for *all*.

Reference List

Behm-Morawitz, E. (2014). Examining the intersection of race and gender in video game advertising. *Journal of Marketing Communications*, *23*(3), 220–239. https://doi.org/10.1080/13527266.2014.914562

Browning, K. (2021, July 22). Activision Blizzard Is Sued by California Over Workplace Culture. *The New York Times*. https://www.nytimes.com/2021/07/21/business/activision-blizzard-california-lawsuit.html

Buyukozturk, B., Gaulden, S., & Dowd-Arrow, B. (2018). Contestation on Reddit, Gamergate, and movement barriers. *Social Movement Studies*, *17*(5), 592–609. https://doi.org/10.1080/14742837.2018.1483227

Cook, J. (2014, October 20). *Twitch Founder: We Turned A "Terrible Idea" Into A Billion-Dollar Company*. Business Insider. https://www.businessinsider.com/the-story-of-video-game-streaming-site-twitch-2014-10

Dietz, T. L. (1998). An Examination of Violence and Gender Role Portrayals in Video Games: Implications for Gender Socialization and Aggressive Behavior. *Sex Roles*, *38*(5/6), 425–442.

eMarketer Editors. (2020, September 3). *US Twitch Usage Accelerates amid Lockdowns*. Insider Intelligence. https://www.emarketer.com/content/us-twitch-usage-accelerates-amid-lockdowns

Entertainment Software Association. (2021). *2021 Essential Facts About the Video Game Industry*. Entertainment Software Association. https://www.theesa.com/resource/2021-essential-facts-about-the-video-game-industry/

Farrell, T., Novotny, J., Fernandez, M., & Alani, H. (2019). Exploring Misogyny across the Manosphere in Reddit. *Conference on Web Science*.

Fox, J., & Potocki, B. (2015). Lifetime Video Game Consumption, Interpersonal Aggression, Hostile Sexism, and Rape Myth Acceptance. *Journal of Interpersonal Violence*, *31*(10), 1912–1931. https://doi.org/10.1177/0886260515570747

Fox, J., Ralston, R. A., Cooper, C. K., & Jones, K. A. (2014). Sexualized Avatars Lead to Women's Self-Objectification and Acceptance of Rape Myths. *Psychology of Women Quarterly*, *39*(3), 349–362. https://doi.org/10.1177/0361684314553578

Gardner, M. (2021, March 19). *The Ugly Hatred Faced By Women In Esports, And How We Fight Back*. Forbes. https://www.forbes.com/sites/mattgardner1/2021/03/19/the-ugly-hatred-faced-by-women-in-esports-and-how-we-fight-back/?sh=5117bc7b37d8

Gestos, M., Smith-Merry, J., & Campbell, A. (2018). Representation of Women in Video Games: A Systematic Review of Literature in Consideration of Adult Female

Wellbeing. *Cyberpsychology, Behavior, and Social Networking*, *21*(9), 535–541. https://doi.org/10.1089/cyber.2017.0376

International Game Developers Association. (2021). *Developer Satisfaction Survey 2021 Report on the Impact of COVID-19*. https://igda-website.s3.us-east-2.amazonaws.com/wp-content/uploads/2021/07/31184838/IGDA-DSS-2021-COVID-Report_July-18-2021-1.pdf

Kondrat, X. (2015). Gender and video games: how is female gender generally represented in various genres of video games? *JOURNAL of COMPARATIVE RESEARCH in ANTHROPOLOGY and SOCIOLOGY*, *6*(1), 188. http://compaso.eu/wpd/wp-content/uploads/2015/08/Compaso2015-61-Kondrat.pdf

Kuznekoff, J. H., & Rose, L. M. (2012). Communication in multiplayer gaming: Examining player responses to gender cues. *New Media & Society*, *15*(4), 541–556. https://doi.org/10.1177/1461444812458271

Limbong, A. (2022, March 29). Judge approves Activision Blizzard $18 million settlement in sexual harassment suit. *NPR*. https://www.npr.org/2022/03/29/1089577389/judge-activision-blizzard-settlement-sexual-harassment

Lorenz, T., & Browning, K. (2020, June 23). Dozens of Women in Gaming Speak Out About Sexism and Harassment. *The New York Times*. https://www.nytimes.com/2020/06/23/style/women-gaming-streaming-harassment-sexism-twitch.html

Massanari, A. (2016). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, *19*(3), 329–346. https://doi.org/10.1177/1461444815608807

Nakandala, S., Ciampaglia, G. L., Mokoto Su, N., & Ahn, Y.-Y. (2016). *Gendered Conversation in a Social Game-Streaming Platform*.

Paaßen, B., Morgenroth, T., & Stratemeyer, M. (2016). What is a True Gamer? The Male Gamer Stereotype and the Marginalization of Women in Video Game Culture. *Sex Roles*, *76*(7-8), 421–435. https://doi.org/10.1007/s11199-016-0678-y

Ruvalcaba, O., Shulze, J., Kim, A., Berzenski, S. R., & Otten, M. P. (2018). Women's Experiences in eSports: Gendered Differences in Peer and Spectator Feedback During Competitive Video Game Play. *Journal of Sport and Social Issues*, *42*(4), 295–311. https://doi.org/10.1177/0193723518773287

Scharrer, E. (2004). Virtual Violence: Gender and Aggression in Video Game Advertisements. *Mass Communication and Society*, *7*(4), 393–412. https://doi.org/10.1207/s15327825mcs0704_2

Takahashi, D. (2021, March 9). *Newzoo: Game livestreaming should grow 10% to 728.8M viewers this year*. VentureBeat.

https://venturebeat.com/2021/03/09/newzoo-game-livestreaming-will-grow-10-to-728-8m-viewers-this-year/

Twitch.tv. (n.d.). *Moderation on Twitch*. Www.twitch.tv. Retrieved April 11, 2022, from https://www.twitch.tv/creatorcamp/en/setting-up-your-stream/moderation-and-safety/

Appendix