

## **Don't be an ASS: Examining Rude Language Behaviors in Games**

### **Abstract**

In this study, we examine, quantify, and contextualize rude speech used in games during gameplay. We identify taboo utterances and use rating categories in an attempt to identify “toxic” behavior in online games. Our study applied a quantitative content analysis to an existing corpus of gaming communications texts (forums, emails, chat and voice) collected during a 6-week study. We define face-threatening texts as “toxic” and found that over half of the identified utterances with curse words were toxic. We conclude that face-threatening utterances contribute to the perceived negativity of gaming environments. We propose that further study investigate texts for implicit toxic language and examine the “tabooness” of certain words at use in-game conversations.

**Keywords**— Toxic language, gaming culture, taboo words, incivility, online disinhibition

### **Background**

In 2015, Zoë Quinn, a game developer and writer, was subject to attacks after posts were made from a jilted lover in the 4Chan and Reddit forums. She also received death threats, was stalked, and nude photos of her were posted. Quinn’s incident precipitated #GamerGate (Hathaway, 2014)—an all-encompassing term that represents either the aforementioned incident, the perpetrator’s actions in the incident, or the ad hoc cultural movement which objects to feminism in the gaming world. These recent and ongoing events of behavior in games has led to the perception that online gaming is toxic, characterized by bullying, cursing, and violence.

According to Kwak and Blackburn, toxic behavior is defined as, “bad behavior that violates social norms, inflicts misery, [and] continues to cause harm after it occurs and affects the entire community”(Kwak & Blackburn, 2014, p. 209). With regards to language, a subset of behavior, the research seems to support the reported experience that gaming culture is also rife with misogynistic and toxic speech (Johnson, 2014). Incidents such as #Gamergate provide a series of compelling anecdotal accounts that bolster this claim. However, while these anecdotal reports can provide a significant launching pad for systematic academic inquiry into a phenomenon, to date, there is as yet, any study that systematically and quantitatively examines the language used in gameplay. That leaves us with the obvious question, beyond the anecdotal accounts, “are gaming environments toxic?”

This research has three broad applications. First, if a baseline can be established for toxic language in games, then gaming gatekeepers (designers, developers, producers, and administrators) can use this data to help find, assess, curtail, and possibly even predict and prevent, bad behavior in game. As a business practice, this just makes good common sense as women make up 47% of the market share of all video games (Womens Media Center, 2014). Second, the newly evolving “corporate citizenship” roles of industry stakeholders dictate that gatekeepers exercise societal governance and address even the perception of unfairness as these issues impact the industry at large (Busch, Boudreau, & Consalvo, 2015, p. 176). Third, for the academy, gaming environments are ripe for language studies because they are digitally

constructed, permit the capture of the linguistic interaction of people in the environment, particularly naturally-occurring emotional speech, enabling us to examine behavior and language use.

In this study, we examine, quantify, and contextualize toxic speech in games using naturally-occurring emotional dialogue captured during gameplay. In keeping with Kwak and Blackburn's (2014, p. 209) assessment that "linguistic components are a prime method of expressing toxicity," we use the corpus of oral and written communication produced by twenty-four participants who played World of Warcraft (WoW) in two teams, over a five-week period. We analyze the context of toxic texts—described by Mehlenbacher as all written, oral, electronic forms of communication (2013)—in an effort to further the academy's understanding of toxic language use in online games. We caution readers that examples used come from the data collected from the study and may be considered offensive.

## Literature Review

### Defining Toxic Behavior

There are numerous terms which can be applied to the phenomenon of toxic behavior. Kwak and Blackburn define toxic behavior as "bad behavior that violates social norms, inflicts misery, continues to cause harm after it occurs and affects the entire community" (2014, p. 209). This bad behavior can include cyberbullying (Ballard & Welch, 2015) and grieving (Chesney, Chuah, & Hoffmann, 2009), typically caused by what Suler (2004), and others (Sproull:1986uj; Dubrovsky, Kiesler, & Sethna, 1991) coin as online disinhibition—the lowering or lack of restraint of social conventions and etiquette and while online, some people act out "more frequently or intensely" than they would in person (Suler, 2004, p. 321). Many agree, however, that toxic behavior is difficult to define because norms differ across games and individuals (Chesney:2009dl; Foo & Koivisto, 2004). That is to say, behaviors such as cursing or grieving may be acceptable in one gaming environment but not acceptable in another.

### Defining Incivility

Research on incivility provides us with a lens to understand and categorize toxic behavior. The context in which incivility occurs is expanding, becoming increasingly conspicuous. Reports continue to arise of uncivil behavior in public spaces including the political arena, the classroom, places of entertainment, and the marketplace. As with bad behavior in public spaces, researchers like Sherry Turkle (2015) have become increasingly concerned about incivility in online spaces as well.

With regards to online spaces, games are becoming the frontrunner in the bad behavior race as evidenced by recent events such as #Gamergate (Higgins, 2015), threats against prominent women in the computer industry (Wingfield, 2014), and other misogynistic instances affiliated

with gaming (R. Greenfield, 2013). One theory advanced by Blackburn (2014) suggests that the gaming environment is particularly vulnerable to this bad behavior because Computer Mediated Communication (CMC) without face-to-face interaction can lead to hostility and aggression. Furthermore, some games due to their goal-oriented nature, like sports, are particularly vulnerable to uncivil behavior (Coe, Kenski, & Rains, 2014). Most of the literature points to the toxic problem being especially prevalent in competitive gaming environments (Märtens, Shen, Iosup, & Kuipers, 2015).

Lately, games and gaming environments have borne the brunt of calls for curtailing uncivil behaviors. Perhaps because gameplay and behavior are reviewable through recordings, chat, and other logs, game environments are the logical place to begin. Most Massively Multiplayer Online (MMO) games have had to make decisions about how to manage player behavior. Terms such as “corpse camping”, “griefing”, and League of Legends’ “player feeding” to name just a few, have been identified and tied to reporting methodologies. Bad behavior is not only difficult to define but its characterization is inherently subjective; that is to say, there is no universal agreement on bad behavior. While not perfect, a more precise measure of toxic behavior may be established through a linguistic framework, since there are well-established semantic and lexical approaches to language use.

### **The Relationship of Toxicity and Swearing**

One of the commonly examined forms of uncivil language is swearing. Jay and Janschewitz (2008) define swearing as the “use of taboo language with the purpose of expressing the speaker’s emotional state and communicating that information to listeners.” These researchers posit that swear words or taboo words can be uttered in any emotional state of the speaker (e.g. joy, sadness, anger, etc.). Additionally, they argue that propositional swearing can be polite or impolite or potentially neither. It is polite when it promotes harmony as in face building (e.g., that dress is pretty fucking hot!); it is rude when used to deliberately attack someone as in face threatening (e.g., You fucking asshole!). This distinction is similar to Hughe’s (2015) three modes of swearing: the expletive or exclamation (damn!), the curse (damn you), and the intensive (damn shame) (p. 170). In essence, not ALL swearing is toxic, especially since swearing, which happens frequently in public, is often conversational, and not highly emotional, confrontational, rude or aggressive (Jay & Janschewitz, 2008). Incidents of rudeness are on the rise, not just in gaming which we explore in this article, but also in society (Wallace, 2017).

However, Jay and Janschewitz (2008) do argue, that when swearing is perceived as an affront to listeners, the speech may be considered rude or uncivil on behalf of the speaker. There are several factors that influence the listener’s perception of rude speech, these include participants’ identity, relationship, social norms, intentions, and motivations. More specifically, Jay and Janschewitz (2008) point to contextual variables which include the topic of conversation, the speaker-listener relationship, gender, occupation, status, the social physical setting (e.g. public vs. private) one’s jurisdiction over the location, and level of formality of the occasion.

In the context of our own work, we agree with Jay and Janschewitz (2008) that not all swearing is problematic, and that there are several factors to consider in the determination of whether swear words are offensive. Additionally, we agree that it is a difficult judgment to make given the fact that the perception of offensiveness is determined by so many interacting factors. However, even in the context of gaming, where we would expect to see higher occurrences of swearing because of the informal environment, we argue that swearing that is face threatening (FT) is offensive and may contribute to the perception of a toxic environment. It is important to note that we use face threatening here as a way of describing the communication text and we are not trying to invoke more formal, theorized notions of face-threatening acts per se. We return to Jay and Janschewitz's (2008) definition that swearing involves the use of taboo words and in studies about toxic language, taboo words are often used as a starting point because according to Jay (2009) almost always taboo words are situationally inappropriate as well as easy to identify (David, 2002). From this point forward and to simplify the number of terms, we will refer to swear words as taboo words.

**Taboo words used in games.** Kwak and Blackburn (2014) used linguistic analysis to compare toxic and typical language behaviors in games. In their study, they used crowdsourced data from nearly 600,000 reports of toxic behavior in League of Legends (LOL). Using n-gram analysis, they identified words used by toxic players compared to typical players. They categorized toxic players based on the reports received through LOL's tribunal (for reporting bad behavior).

While the study has some very interesting findings, for us, what is most useful is that the researchers (Kwak & Blackburn, 2014) are able to establish a baseline of the most commonly used words by players, which we use in our study. According to Kwak and Blackburn (2014) for toxic players, these words included: "retards", "nigger garbage", "uninstall", "fuckign [sic]", "bots", "fucking retard", "report noob", "fuck team", "stupid noob", "pussy ass", and "play fucking". They astutely note that some of these words, like "fucking" aren't used exclusively by toxic players and as we mentioned previously may only be perceived as egregious in certain circumstances.

The study defines typical players as a "set of players on the opposite team when none of them report the toxic player." Kwak and Blackburn's (2014) use of the idea of the "opposing team" to define "typicalness" suggests that players on the same team would not report a teammate, even if the teammate used toxic speech, and consider that speech, normal conversation between players on the same team. We see this differently and argue that even when toxic language occurs *within* the team, it contributes to the toxic culture of a game. Additionally, researchers inadvertently discount the significance of acts that are face threatening, by measuring toxicity only when aimed at the competitor (a common binary contrivance in gaming) but discounting any issues toward members of your own team. For us, this study points to the further need to investigate and examine instances of FT in games; a key methodological distinction, which we make in our study. Therefore, we examine taboo words for their frequency of occurrence and usage.

However, we stress that it is not simply the use of taboo words that are at issue, but instead it is the context of their use.

### **Research Questions**

Given the concerns about toxic behavior and rudeness, we have two overarching research questions:

1. What percentage of gamer texts contains taboo utterances?
2. How rude and how toxic is gamer text?

For the first research question, we are interested in establishing a baseline of taboo behavior in games, using the utterance as the unit of analysis (discussed further in our Methods section). For the second research question, we assert that we can uncover opportunities for rudeness and toxicity by searching for and identifying taboo words. We then use raters to differentiate between rudeness and toxicity in taboo-laden texts. In the following methods section, we include an overview of the study, describe the participants and raters, enumerate our hypotheses, discuss our choice of unit of analysis, identify our measures, and describe our materials and protocol.

### **Research Methodology**

This study examines texts collected from a previous gaming study (Robinson, 2014) on leadership. In this study, remote gamers played World of Warcraft's end-game raids approximately two times per week for about four hours each time, accumulating over 100hrs of text. During this period, gamers used communication technology for a variety of tasks, including discussing the challenges of the game, planning for future sessions, debating strategies and gaming process, and sometimes to disparage, console, or make friends. The gamers communicated through chat, voice, forums, and email. Chat was captured from the game's interface and was only available while gamers were logged into the game. Voice was captured using separate push-to-talk software called Ventrilo; it was started at the beginning of each gaming session and ran concurrently while players were logged into the game. Email and forums were available at any time during the five weeks and were mostly used outside of gameplay. This entire corpus was recorded/captured/downloaded and later transcribed. For this study, the reuse of this de-identified corpus received IRB approval in July 2015 (IRB# 2012-061) and consisted of 1,950 utterances (or 26,519 words).

Our goal was to have raters examine the data corpus and apply a code to an utterance of text. In order to create a corpus for raters to examine, we began with the list of taboo words from Kwak and Blackburn's (2014) study. To this list, we added a select few words (e.g., synonyms) resulting in a list of 26 taboo words (see Appendix A). Using Nvivo, we searched the corpus for the taboo words and pulled the text containing the taboo word as well as the surrounding few

words into a separate file. These passages were further scrubbed of all avatar names, replacing them with “speaker 1”, “speaker 2” nomenclature, and the taboo words were emphasized in red. This process reduced the sample to 327 utterances that required coding. Identifying the taboo word in red alleviated the burden on the raters to search for and locate the taboo words and instead permitted them to concentrate on evaluating the context and applying the most appropriate codes.

The twenty-four gamers in the study were recruited from forums and social media sites. The participants’ ages ranged from 18 to 45 and the majority of participants were Caucasian (76%), male (64%), and lived in the US (92%) across various cities. More than 50% had some college education and on average over 6 years of work experience as well as 5-years experience of playing WoW. Most of the participants worked full time (36%) and 28% were still attending college. Our three raters were female and college graduates who ranged between the ages of 30-45; two identified themselves as gamers. One of the participants was still attending college and two were full-time academics.

Using the codebook (see Appendix B), the three raters initially coded 15 pages of text (approximately 48 utterances). Coders then had a discussion to clear up any discrepancies on the text and codes. The initial coded sample (after group discussion) was rated a  $\kappa=0.85$ , using Cohen's kappa which “provides an assessment of the level of agreement between observers”, where  $\kappa$  greater than 0.61 indicates substantial agreement (Warner, 2013). Subsequently, raters coded the remaining 115 pages and the final interrater reliability was  $\kappa= 0.78$ . The data was then analyzed using simple descriptive statistics namely, frequency counts.

The utterance with the taboo word highlighted in red and its surrounding text was provided as a series of rows collected into an electronic Word file. The amount of surrounding text varied by medium, but in most cases, it was a series of sentences prior to the utterance and a series of sentences after the utterance. Raters were tasked with reviewing any highlighted words using the surrounding text as context and applying a single code that would categorize the taboo utterance. Seven codes, based on an a priori codebook (see Appendix B), were available. Utterances were either classified as rude (comments that contained taboo words but were not face threatening) or toxic (comments that contained taboo words that were face threatening). The toxic texts were classified further by whether it was directed at a person or group (i.e., name calling, aspersion, pejorative, and critical). All utterances containing an identified taboo word were coded by only one code; raters needed to apply the single code that fit the utterance best. Utterances were also categorized by synchronicity—synchronous texts were voice and chat and asynchronous were forums and email.

### **Unit of Analysis**

The study is one of the first to categorize overt rude language usage by coding the author’s/speaker’s words. To accomplish this, we used the utterance as the unit of analysis. The utterance size—the smallest unit of coding—was specified by the original study (Robinson, 2014). More than a simple turn, an utterance is the talk/conversation delivered to one recipient.

The corresponding reply/response to the sender is another utterance. Thus, an utterance is predicated on a single sender and recipient pair, regardless of length. In the case of voice, it is when the speaker stops speaking, but in the case of email, it is the entire email.

The use of utterances helps to better contextualize the comments and to provide a method of comparison across media—voice, chat, email, and forums—(Heckman & Misiolek, 2005). The categorization of single words would lead to ambiguous results. For example, an expletive like “shit” can be a synonym for excrement as “I stepped in shit”; an interjection used for emphasis as “Shit! I missed the bus”; or name calling as in “You’re a piece of shit.” While all three utterances contain taboo words, it is still difficult to assess whether any one of the utterances is rude or toxic, without the surrounding text. However, with the additional context the text becomes more clear. For example, take the latter utterance, “You’re a piece of shit.” On the face of it, it looks like this utterance could be toxic to the listener since it is directed at a person. However, if the entire passage read, “I can’t believe you got me this ring! I had no idea, and we were just at the jewelers yesterday. *You’re piece of shit!* I love it!”, this utterance would not be coded as toxic. It is *not* an attack on the hearer; therefore, it would be coded as “rude” for simply containing the taboo word “shit”. In this study, we used the utterances and the surrounding words to code the corpus.

## Measures

Raters were directed to code every utterance. Following Coe (2014), we used their forms of incivility from their study on political discourse<sup>1</sup>. Appendix B shows the full codebook including the codes for unclear utterances. We want to stress that coding the utterances enabled us to determine which of the previously identified taboo utterances were perceived as toxic. In essence, the raters were coding whether taboo words such as “fuck” were used in a manner that might contribute to the perception of toxicity because it was an apparent attack on the listener. Further, the codes allowed us to understand the types of rude utterances in gameplay.

## Hypotheses

Our research questions state that we are interested in both the frequency of taboo utterances in gaming as well as the perception of toxicity in the gaming environment. With regards to our first research question, given that no prior study using gaming texts has used utterances as a unit of analysis, we have no basis for developing a hypothesis to examine this question. We will simply report on the findings.

For our second research question, we define “rude texts” as any texts that contain curse words and “toxic texts” as any that have identified as face threatening or FT. We hypothesize that if the

---

<sup>1</sup> With the exception of lying (since the act of calling out a lie doesn’t have to explicitly use taboo words).

gaming environment is toxic, as has been suggested by anecdotal reports, we would expect instances of toxic texts to occur more frequently than rude texts. Also, with regards to gameplay, since communication happens within and across different media (e.g., chat or email), whether the medium is synchronous or asynchronous should also impact the perception of toxicity in the environment. That is to say, when communication happens in real time, we argue that it is perceived differently than when it is left for the observer to review later. Words or utterances spoken where the impact can be felt immediately rather than mulled over outside of any interaction would seemingly be important to review. Therefore, we hypothesize if the gaming environment is perceived to be toxic, then toxic utterances will occur more in synchronous media (i.e., chat) than in asynchronous media (i.e., email). In order to measure toxicity in the gaming environment, we observe the frequency of occurrence of coded utterances in each medium and across synchronicity (i.e., synchronous or asynchronous ). To summarize:

**H1:** Toxic texts will occur more frequently than rude texts.

**H2:** Toxic texts will occur more frequently in synchronous media than asynchronous media.

## Results and Discussion

### RQ1: What Percentage of Gamer Texts Contains Taboo Utterances?

Table 1 shows that 17.2% of the corpus contains taboo utterances with forums and email having the highest percent of taboo utterances at 100% and 40% respectively. It is worth noting that the volume of email and forums communication produced during the study was low, 15 and 10 utterances respectively; most participants communicated synchronously through chat and voice.

Table 1. Number of taboo utterances in the corpus

	<b>Number of Taboo Utterances</b>	<b>Total Utterances</b>	<b>% Taboo Utterances</b>
<b>Email</b>	6	15	40.0%
<b>Forums</b>	10	10	100.0%
<b>Chat</b>	54	322	16.8%
<b>Voice</b>	257	1,558	16.5%
<b>Total</b>	327	1,905	17.2%

## RQ2: How rude and how toxic is gamer text?

Our first hypothesis states that “toxic texts will occur more frequently than rude texts.” A breakdown of types (see Figure 1) shows for the coded taboo texts, the majority of utterances (54%) were classified as toxic texts; inside this classification, “critical of performance” was the highest category (24%) and name calling next (21%). Even though toxic texts show a slim majority, this result supports our hypothesis, and our assertion that the perception of toxicity might be predicated on a personal attack and not simply the frequency of taboo language use.

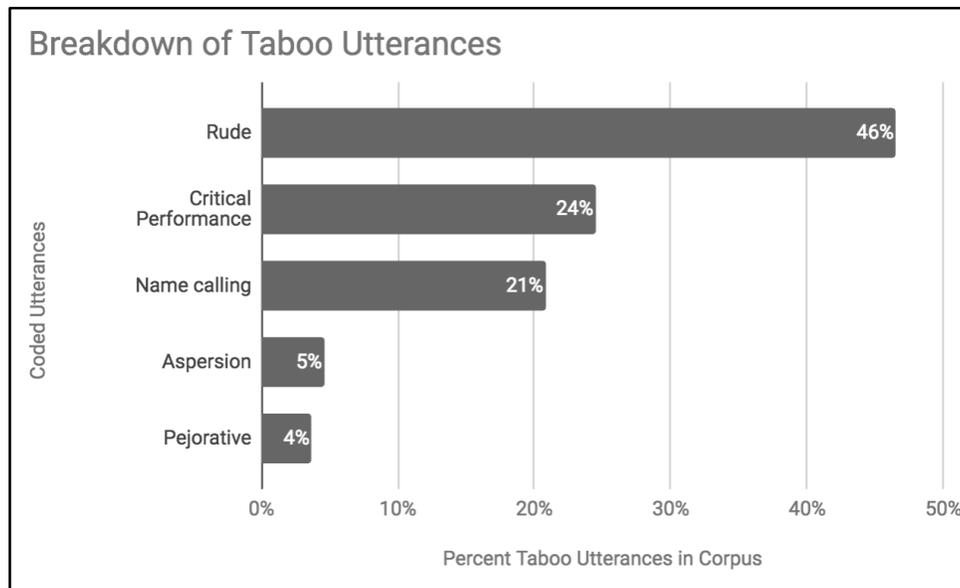


Figure 1: Coding of Utterances in the Corpus

Our second hypothesis asserts, “toxic texts will occur more frequently in synchronous media than asynchronous media.” The medium with the most taboo content was the forums (100%). The forums also had the majority of toxic texts at 40% (see Table 2) with email containing only half that number (20%). Chat and voice contained 8.4% and 9.1% respectively of the toxic texts.

Table 2: Percent of Toxic Utterances by Medium\*

Utterance Level Analysis						
	Number of Rude Utterances	% Rude Utterances	# of Toxic Utterances	% Toxic Utterances	Total Utterances	% Taboo Utterances
Email	3	20.0%	3	20.0%	15	40.0%
Forums	6	60.0%	4	40.0%	10	100.0%
Chat	27	8.4%	27	8.4%	322	16.8%
Voice	116	7.4%	141	9.1%	1,558	16.5%
<b>Total</b>	152	8.0%	175	9.2%	1,905	17.2%

\*The shaded rows indicate the asynchronous mediums.

We combined email and forums and chat and voice to examine toxic texts across synchronicity or the distinction between asynchronous and synchronous communication (see Figure 2).

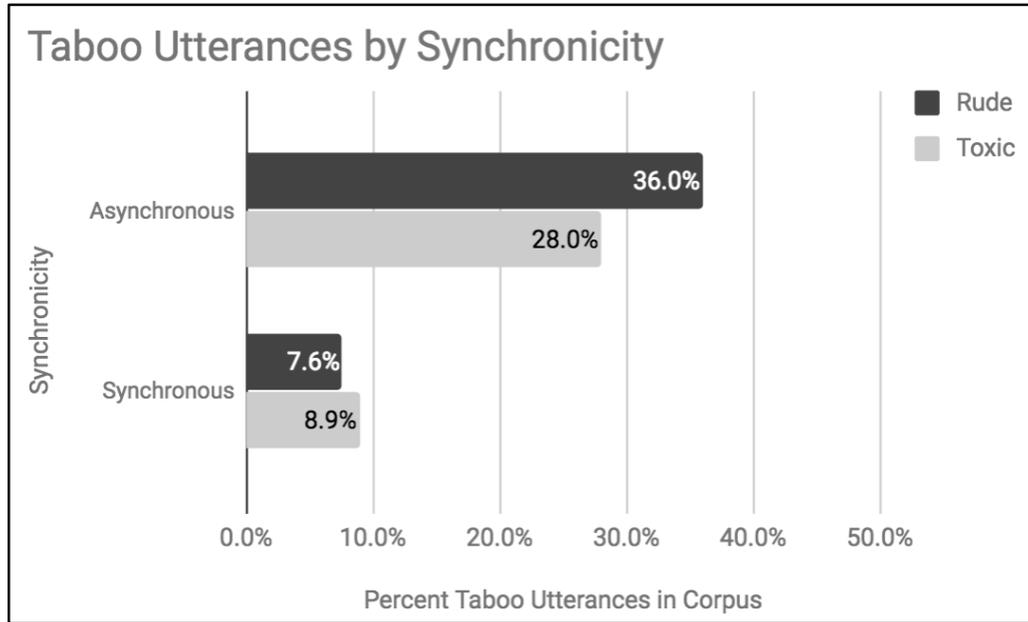


Figure 2: Percent rude and toxic utterances

For asynchronous communication, the frequency of rude texts was 36% and for toxic texts it was 28%. For synchronous communication, the frequency of rude texts was 7.6% and for toxic 8.9%. Therefore, based on these results, Hypothesis 2 is not supported. Table 3 summarizes our hypotheses.

Table 3. Hypotheses using Utterance Level Analysis

Hypotheses	Supported?
<b>H1:</b> Toxic texts will occur more frequently than rude texts.	Yes
<b>H2:</b> Toxic texts will occur more frequently in synchronous media than asynchronous media.	No

## Discussion

In thinking about our first research question “What percentage of gamer texts contains rude utterances?”, our data shows 17% of the total corpus has rude utterances. That in of itself seems relatively high; almost 1 in every 5 messages has some type of uncivil text. If we take a deeper look into the synchronous texts, chat rude text utterances are at 16.8%, for example (refer Table 1) Unfortunately, we have limited studies with which to compare our results and no researchers have coded and examined the utterances of texts in MMO games. Therefore, we offer our findings as a baseline of taboo language usage in this domain.

In relation to the second research question “How toxic is gamer text?”, we have two hypotheses to discuss. For the H1, we found that toxic texts occurred more frequently than rude texts. Therefore, we reassert that it is the relatively higher percentage of toxic texts that contribute to the poor perception of the gaming environment. That is to say, in the course of gameplay, players are confronted with comments that criticize them, attack their ideas or speech, and/or are critical of their play; players face more toxic comments than acts of rudeness.

H2 stated that “Toxic texts will occur more frequently in synchronous media than asynchronous media.” Our results found that H2 was unsupported; forums and email have more utterances that were coded as toxic than chat or voice. However, we posit it is the comparison across modes that gives us a poor understanding of behavior in this system. This claim is predicated on the inadequacy of an utterance-level analysis in comparisons *across* asynchronous and synchronous media. This inadequacy arises from the definition of an utterance—defined as a single sender/recipient turn *regardless* of the length of the communication. One example may help to clarify this issue. When compared to an email or forum post, text produced in a single turn in chat or voice is often shorter, characterized by fewer words in the utterance. Therefore, if a forum has 20 words across 1 utterance and a chat has 5 words across 4 utterances, and each has only one taboo word in 20 words, then 5% of the forum text is taboo compared to 20% of chat text. This is a definitional issue and it creates disproportionate “tabooness” for asynchronous media. However, we assert it is possible to compare synchronous modes to each other (for example, voice with chat) and asynchronous modes to each other (forums to email). In the following section, we discuss studies that show these comparisons.

**Asynchronous toxic studies.** Asynchronous texts are typically high in toxicity (Groshek & Cutino, 2016; Santana, 2013). Santana (2013) used newspaper comments (similar to forum posts) and compared anonymous to non-anonymous newspaper comments about immigration in three major newspapers. A comment was rated as toxic when “at least one of the following was present: personal or inflammatory attacks, threats, vulgarities, abusive or foul language, xenophobic or other hateful language or expressions, epithets or ethnic slurs, sentiments that are racist or bigoted, disparaging on the basis of race/ethnicity or that assign stereotypes” (Santana, 2013, p. 25). Santana found 41% toxic comments in aggregate. In a similar, more recent study on Twitter data, researchers coded a sample of approximately 2,700 captured tweets that provided a

distinct but “diverse overview of contentious politics” (Groshek & Cutino, 2016, p. 3) and found that 36.1% were toxic. Our combined asynchronous results are lower than either of these studies at 28.0% toxic texts usage (refer Figure 2); however, looking at forums (characteristically more similar to Twitter and comments) it was 40.0% (refer Table 2). We should note that our asynchronous data is sparse, a mere 25 total asynchronous utterances; too few data points to make a statement about the significance or meaning of the difference.

**Synchronous taboo studies.** These types of communications are the most prevalent in MMORPG arenas and therefore require additional examination. Our study showed that ~9% of taboo utterances in synchronous media were toxic (refer Figure 2). It is relatively startling that approximately 1 of every 6 gamer utterances is at best, uncivil and at worst, 1 out every 10 is toxic. With regards to prior research, only one study examined chat using utterances. Subrahmanyam (2006) found that 5% of utterances from unsupervised chat rooms contained bad language; our results of taboo utterances usage in chat is considerably higher at 16.8% or more than 3 times that of Subrahmanyam’s results (refer Table 2). We hold the view that our synchronous data is elevated due to the nature of MMORPG gaming. Much like Coe et al. (2014) and Martens et al. (2015), we posit that the competitive nature of gaming, which is both emotional and goal-directed, provides a distinctly different environment in which potentially toxic language is produced when compared to the Subrahmanyam (2006) study. It is the innate need to win/succeed against either the game or other teams that shapes both the production of taboo words and the deployment of toxic texts.

**Effect on players.** Our results raise interesting implications about what effect these levels of toxicity have on players. Our results show that the gaming environment contains ~17% taboo texts and 9% toxic texts which disparage, or otherwise abuse the listener. These facts taken together arguably make for an environment that is antithetical to our traditional concepts of play. We propose a language-driven approach to address the gaming environment. Using linguistic markers, we advocate for practices that: 1) establish a clear set of guidelines or “etiquette” for language use in gaming environments that can point to concrete examples of offensive, face threatening language, 2) focus attention on certain media (i.e., chat) as a likely source of toxic behavior over other media; 3) develop tools that use language to help both players self monitor and, hopefully, curb their toxic behavior, and game masters censure repeat toxic offenders, and/or 4) provide opportunities for players to opt-out of or mute toxic conversations. We recognize that some of these ideas, while not new, can be now be implemented as described below given the sophistication of AI and voice to text algorithms available today (Carmiel, 2017).

First, developing a set of clear guidelines for players to use during gameplay might help reduce toxicity in the environment. These guidelines might be shown to players on a rotating and random basis at different junctures during gameplay. These contextual reminders might go a long way to creating a more civil gaming environment. Secondly, the collection and identification of utterances through by machine language algorithms could provide an opportunity to predict toxic

behavior. Blackburn and Kwak advanced that with enough data, the prediction of events that are most likely to precipitate toxic behavior is possible (2014). Third, since prediction is not currently possible, using the aforementioned categories for toxic behavior, a toxicity meter might be developed to help “measure” and indicate the amount of toxicity found in the language of a speaker/player. This type of meter could be easily coupled with chat conversations and presented in a head-up display for the player. This meter would provide feedback to the player in an effort to promote self-policing of toxic behavior. Teams that have consistently low toxicity meter readings could be provided with incentives to maintain this behavior in the game. This meter could also be tied directly to player performance and as a player’s rating worsens, the player’s stats would decrease as well. Simultaneously, players with toxicity meters that are consistently and repeatedly in “red” toxicity levels might have their cases escalated to game masters. Official censure would result in repeat offenders privileges being either curtailed or revoked. Lastly, allowing players to mute or eject players that have repeated violations could be an extension of the toxicity-metered approach. Players in teams that have consistently “red” toxicity levels might trigger a vote by other teammates to either mute or eject the offending player from the team. Players who are repeatedly admonished for bad behavior may eventually be expelled from the game.

## Conclusions

One of the major limitations of the study is that because gaming language is difficult to understand, there is an inherent challenge of choosing raters for this type of study. One obvious issue in asking raters to make judgments about language use and participant intention is that gaming language can be very exclusive — a sort of gamespeak if you will, with its own jargon. Consequently, it raises the question about who are the best raters for a study like this. On the one hand, asking gamers to rate speech may be inviting a type of expert “bias” into a study as other gamers may be oblivious to toxic language use because of their comfort or familiarity with the environment and language used in that environment. On the other hand, non-gamers may be overly sensitive to the toxic language used in games (Kwak & Han, 2015) or depending on their natural environment, under sensitive.

Additionally, this study used data from WoW—an MMORPG. Given that different environments are perceived to have varying levels of toxicity, for example, LOL’s notoriety for toxicity is renowned (Maher, 2016), it is no stretch to believe our study repeated in different gaming environments would yield different results. Hence, we caution against the generalization of this data to all games and gaming environments, given that the variations in toxicity might be mitigated or exacerbated by social game structures such as guilds, or raid groups or existing genre contrivances such as avatar race, gender, position in society, or game goals (e.g., capture the flag, battlegrounds, or tournaments). Our study is a glimpse into the MMO environment and not a comprehensive look at the gaming industry; therefore, our results may not be applicable to other contexts.

### **Observance of Naturally-occurring Language**

This study is unique because the texts collected and analyzed contained naturally-occurring, emotional speech. Data on emotional speech typically have been collected in one of two forms: performed/acted and naturally-occurring speech. Both forms are subject to considerable debate and face what the literature calls the challenge of authenticity (Cullen, Vaughan, Kousidis, & Wang, 2006). On one hand, while acted speech appears in a controlled setting, the method of collection is problematic because it is unclear to what extent acted speech is representative of real emotions (Cullen et al., 2006). On the other hand, while naturally-occurring speech is ideal for "high ecological validity" (Scherer, 2003, p. 232), fully spontaneous emotional speech is also problematic because it typically does not appear in a controlled setting. Our ability to collect a high-quality recording of naturally-occurring, emotional speech in quasi-controlled conditions suggests that speech produced during real games is a reliable data source and lays the groundwork for certain types of language variation studies.

### **Multi-media study and its affordances**

Quite often, media and language studies, focus on communication generated in a single medium with the same set of individuals (Kwak:2014ig; Coe:2014iz; Santana:2013cn; Richter, Williams, Magny, & Luechtefeld, 2011). While the advantage of this single-medium approach is clear, in that it allows a thorough investigation of communication within that medium, the reality is that teams communicate using a variety of media to get their tasks done. And since different media have distinctly different affordances, we have an opportunity in a multi-media study to observe participants' use of different media to achieve their communication goals. As a result, and within in the context of this study, we gain a more nuanced understanding of toxic language use in games; for example, we have been able to identify particular media (i.e., chat), where toxic speech is more prevalent. Understanding that some media potentially enable toxic behavior gives us an opportunity to think about the development of methods and tools targeted to that medium.

### **Broader Implications**

Attempting to quantify and characterize the use of toxic language in games is an important first step toward understanding and identifying the nature and scope of incivility in online environments. This study has provided empirical evidence of incivility which corroborates anecdotal accounts about the offensive nature of certain gaming environments. In fact, the use of the utterance as a unit of measurement helps to further contextualize the listeners' experience. Findings like ones in this study, perhaps, can begin to shape the conversation about toxic language in gaming in meaningful ways which include, but are not limited to discussions across contexts, gender, cultures, genres of games played, media and environments, and leadership dimensions. This evidence should provide stakeholders with impetus to move past rudimentary arguments about the gaming culture, and begin to think more seriously about how these environments can be made safer, how toxic players can be sanctioned; and more generally, how

we can create a more inclusive environment for anyone who chooses to participate in online games.

### **Future Research**

Additional empirical research on this topic should consider extending semantic analysis to examine *implicit toxic* language in gameplay. Arguably, implicit toxic language, potentially more insidious, is harder to access, and even harder to contextualize. While not impossible to examine, a very different type of analysis would be required in this case and would likely need to be designed through the theoretical lens of conversational implicature. It is foreseeable that implicit toxic language use is another significant contributor to the perception of gaming's caustic environment.

Lastly, the study raises the question: are some words more toxic than others? We propose the examination of the "tabooness" of the words used in gaming. Perhaps the perception of hostility of gaming environments also results from the severity of taboo word usage. Other studies (Jay:2008ix; Locher & Watts, 2005) contextualize swearing and taboo word usage as dependent on the audience, as well as locale and tabooness. It stands to reason then that different members of the gaming audience might also deem some words more egregious than others. This possibility, coupled with toxic language usage, may have a more intense impact on the hearer and warrants further examination.

## References

- Ballard, M. E., & Welch, K. M. (2015). Virtual Warfare: Cyberbullying and Cyber-Victimization in MMOG Play. *Games and Culture*, 1–27. <http://doi.org/10.1177/1555412015592473>
- Blackburn, J. (2014, September 4). *An Analysis of (Bad) Behavior in Online Video Games*. University of South Florida.
- Blackburn, J., & Kwak, H. (2014). STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games (pp. 877–888). Presented at the Association for the Advancement of Computing in Education, New York, New York, USA: ACM Press. <http://doi.org/10.1145/2566486.2567987>
- Busch, T., Boudreau, K., & Consalvo, M. (2015). Toxic Gamer Culture, Corporate Regulation, and Standards of Behavior Among Players of Online Games, 176–190.
- Carmiel, Y. (2017, August 9). Deep Learning Revolutionizes Conversational AI. Retrieved January 21, 2018, from <https://www.oreilly.com/ideas/deep-learning-revolutionizes-conversational-ai>
- Chesney, T., Chuah, S. H., & Hoffmann, R. (2009). Virtual World Experimentation: an Exploratory Study. *Journal of Economic Behavior & Organization*, 72(1), 618–635.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, 64(4), 658–679. <http://doi.org/10.1111/jcom.12104>
- Cullen, C., Vaughan, B., Kousidis, S., & Wang, Y. (2006). Generation of High Quality Audio Natural Emotional Speech Corpus Using Task Based Mood Induction.
- David, Y. (2002). The Use of Taboo Words by Sangiherese Living at Perak Surabaya. *Surabaya Universitas Kristen Petra*.
- Dubrovsky, V., Kiesler, S., & Sethna, B. (1991). The Equalization Phenomenon: Status Effects in Computer-Mediated and Face-to-Face Decision-Making Groups, 6(2), 119–146.
- Foo, C. Y., & Koivisto, E. (2004). Defining Grief Play in MMORPGs: Player and Developer Perceptions. Presented at the International Conference on Advances in Computer Entertainment Technology.
- Greenfield, R. (2013, June 10). The Rape 'Joke' at Microsoft's E3 Reveal Is a Bigger Deal Than Another Bad 'Joke'. Retrieved July 30, 2015, from
- Groshek, J., & Cutino, C. (2016). Meaner on Mobile (pp. 1–7). Presented at the Proceedings of the 7th 2016 International Conference on Social Media & Society, New York, New York, USA: ACM Press. <http://doi.org/10.1145/2930971.2930976>
- Hathaway, J. (2014, April 10). What Is Gamergate, and Why? An Explainer for Non-Geeks. *Explainer*, 1–5.
- Heckman, R., & Misiolek, N. (2005). Leaders and Followers in Student Online Project Teams (pp. 4c–4c). Presented at the Proceedings of the 38th Hawaii International Conference on System Sciences - 2005.
- Higgins, A. (2015). *Cuties Killing Video Games: Gender Politics and Performance in Indie Game Developer Subculture*.
- Hughes, G. (2015). *An Encyclopedia of Swearing*. Routledge.
- Jay, T. (2009). The Utility and Ubiquity of Taboo Words. *Perspectives on Psychological Science*, 4(2), 153–161. <http://doi.org/10.1111/j.1745-6924.2009.01115.x>

- Jay, T., & Janschewitz, K. (2008). The Pragmatics of Swearing. *Journal of Politeness Research. Language, Behaviour, Culture*, 4(2), 1–23. <http://doi.org/10.1515/JPLR.2008.013>
- Johnson, K. (2014, July 30). *Overt and Inferential Sexist Language in the Video Game Industry*.
- Kwak, H., & Blackburn, J. (2014). Linguistic Analysis of Toxic Behavior in an Online Video Game. In *Social Informatics* (Vol. 8852, pp. 209–217). Cham: Springer International Publishing. [http://doi.org/10.1007/978-3-319-15168-7\\_26](http://doi.org/10.1007/978-3-319-15168-7_26)
- Kwak, H., & Han, S. (2015). Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games (pp. 1–11). Presented at the CHI.
- Locher, M. A., & Watts, R. J. (2005). Politeness Theory and Relational Work. *Journal of Politeness Research. Language, Behaviour, Culture*, 1(1). <http://doi.org/10.1515/jplr.2005.1.1.9>
- Maher, B. (2016). Can a video game company tame toxic behaviour? *Nature* (Vol. 531, pp. 568–571). Nature. <http://doi.org/10.1038/531568a>
- Märtens, M., Shen, S., Iosup, A., & Kuipers, F. (2015). Toxicity Detection in Multiplayer Online Games (pp. 1–6). Presented at the Network and Systems Support for Games, IEEE. <http://doi.org/10.1109/NetGames.2015.7382991>
- Mehlenbacher, B. (2013). What Is the Future of Technical Communication? In J. Johnson-Eilola & S. A. Selber (Eds.), *Solving Problems in Technical Communication* (pp. 1–22). Chicago: University of Chicago Press.
- Richter, N., Williams, L., Magny, O., & Luechtefeld, R. (2011). The Effect of Emotive Language in Virtual Teams (pp. 1–6). Presented at the Frontiers in Education Conference (FIE), 2011, IEEE. <http://doi.org/10.1109/FIE.2011.6142759>
- Robinson, J. (2014, March 26). *Communication, Leadership, and Virtual Teams*. Illinois Institute of Technology.
- Santana, A. D. (2013). Virtuous or Vitriolic. *Journalism Practice*, 8(1), 18–33. <http://doi.org/10.1080/17512786.2013.813194>
- Scherer, K. R. (2003). Vocal Communication of Emotion: a Review of Research Paradigms. *Speech Communication*, 40(1-2), 227–256. [http://doi.org/10.1016/s0167-6393\(02\)00084-5](http://doi.org/10.1016/s0167-6393(02)00084-5)
- Subrahmanyam, K., Smahel, D., & Greenfield, P. (2006). Connecting Developmental Constructions to the Internet: Identity Presentation and Sexual Exploration in Online Teen Chat Rooms. *Developmental Psychology*, 42(3), 395–406. <http://doi.org/10.1037/0012-1649.42.3.395>
- Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, 7(3).
- Turkle, S. (2015, September 26). Stop Googling. Let’s Talk. *New York Times*, pp. 1–8.
- Wallace, J. B. (2017, August 19). The High Costs Of Workplace Rudeness. *Wall Street Journal*, pp. 1–3.
- Warner, R. M. (2013). *Applied Statistics: From Bivariate Through Multivariate Techniques* (2nd ed.). SAGE.
- Wingfield, N. (2014, October 15). Feminist Critics of Video Games Facing Threats in “GamerGate” Campaign. *New York Times*, pp. 1–5. Retrieved from [http://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html?\\_r=0](http://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html?_r=0)
- Womens Media Center. (2014). The Status of Women in the US media 2014, 1–82.

## Appendix

### Appendix A: Dirty Words

The list of dirty little words...(in no particular order):

shit, fuck, shithhead, bitch, cunt, pussy, prick, asshole, ass, twat, stupid, dumb, dick, penis, vagina, pecker, noob, hell, damn, jesus, goddamn, christ, god, sucks, stfu, nigger

### Appendix B: Full Codebook

Codes	Definition	More detail	Examples
VU	rudeness (undirected offensive language)	Offensive language, which include curse or swear words or other low language use, whether in common English or in gamer speak	“I came knowing that people would need help with gear and fights which never makes me feel intrinsically superior or that it gives me some license to disregard social norms like basic decency and treat people like shit.”
NC	Name calling (offensive language directed at the person or group)	Name calling, which includes berating or otherwise demeaning the person	You bitch you are a sorry excuse as a person. I couldn't be happier to not have to deal with you.
AS	Aspersions (offensive language directed at a person's or group's ideas)	The use of offensive language, or a derogatory remark to criticize the ideas or opinions of another.	It's stupid to do it that way! We have got to clear that mob first, otherwise we will be swamped in adds.
PE	Pejorative (offensive language directed at a person's or group's speech or communicative ability)	The use of offensive language: disparaging, derogatory, or belittling remarks to criticize how or what another communicates.	That guy talks like a noob! All caps really?
CR	Critical of performance (offensive language directed at a person's)	The use of offensive language: disparaging, derogatory, or belittling remarks to criticize another's	You're supposed to be taunting those spiders, not pissing in the wind.

	or group's performance)	ability to get a task done.	
NO	No toxic words found. Nothing to code	There was no instance to code	I heard that was bugged right now, so you won't get it noted
U	Unclear	Unable to disambiguate the text. Do not understand.	Alright, amalgamations just spawned shit u alright they're all up, roll right, get inside the wound and we're clear, kill the corruption.