Tea Bagging, Trolling & Trash Talk

A Multimodal Approach for Toxicity in Twitch Streams

Tamara Bodden & Ashhab Ahmed

Abstract Toxicity research already investigates chat and gaming communication. Live streams on platforms like Twitch, however, are a challenging research object due to their multimodal affordances, which enable toxic actions alongside verbal transgressions. Existing approaches that focus only on easily extractable chat communication cannot capture the complex interplay of toxicity among multiple participants in a livestream (players, streamers, viewers in chat) across different multimodal layers. This paper uses a stream (over 94,600 Tokens) of the game *DotA 2* as a basis to describe the different types of multimodal toxicity that can be found in stream and meta discourse, as well as the interplay between the different actors and layers of the stream. The analysis shows that while there is awareness about certain topics that overstep norms (for example, racism), there is a high percentage of actions that can be considered mildly toxic, for example, in using swear words or attacking the streamer in a bantering fashion.

Keywords Toxicity, Stream, Online Gaming, DotA 2, Twitch

Abstract Die Forschung zum Thema Toxizität befasst sich bereits mit Chat- und Gaming-Kommunikation. Live-Streams auf Plattformen wie Twitch stellen jedoch aufgrund ihrer multimodalen Möglichkeiten toxische Handlungen zulassen, die über rein verbale Überschreitungen hinausgehen, ein anspruchsvolles Forschungsobjekt dar. Bestehende Ansätze, die sich ausschließlich auf leicht extrahierbare Chat-Kommunikation konzentrieren, können daher nicht das komplexe Zusammenspiel von Toxizität zwischen mehreren Teilnehmer*innen eines Live-Streams (Spieler*in, Streamer*in, Zuschauer*in im Chat) auf den verschiedenen multimodalen Ebenen erfassen. Dieser Beitrag nutzt einen umfangreichen Stream (über 94.600 Tokens) des

Spiels *DotA 2* als Grundlage, um verschiedene Arten multimodaler Toxizität zu beschreiben, die im Stream- und Metadiskurs zu finden sind, sowie das Zusammenspiel zwischen den verschiedenen Akteur*innen und Ebenen des Streams. In der Analyse wird deutlich, dass zwar grundsätzlich ein hohes Bewusstsein zum Thema Toxizität besteht (und z.B. bestimmte rassistische Äußerungen vermieden werden) aber gleichzeitig zahlreiche Formen von ‚milder' von Toxizität z.B. in der Nutzung von Kraftausdrücken oder in einer Praxis des Beleidigens in Form von ‚bantering' zu finden sind.

Schlagworte Toxizität, Stream, Online Gaming, DotA 2, Twitch

1. Introduction

Twitch is a digital platform primarily used by gamers that reaches over 2 million average concurrent users (Twitchtracker) and attracts an international audience across different age groups. With their interactive mix of 'hot' (live videos) and 'cold' (chat functions) media content (Sjöblom, Hamari 2016: 4), live streams offer a unique communication structure in which meaningful communities emerge and form specific practices (Diwanji et al. 2020: 7). Online multiplayer gaming has always faced high levels of toxicity, i.e. anti-social behaviour in interpersonal communication (e.g. Kwak, Blackburn 2014; Martens et al. 2015), which can also influence streams as an extension of gaming culture. Twitch streams, therefore, can give insights in how to understand (and potentially prevent) toxic behaviour. But they also make a re-evaluation of the classification of toxic actions and speech investigated by prior research necessary that often classifies toxicity exclusively as a negative phenomenon since specific toxic actions seem to be common and accepted practices in many streams. Toxicity is therefore conceptualized following Wittgenstein (1967: PU 23) as *Sprachspiel* (language game) that adheres to specific rules and is embedded in a context (Mall et al. 2020: 2; 1222; Nobata et al. 2016: 152), requiring a pragma-linguistic approach that enables a more adequate evaluation of toxicity. This study, therefore, uses a qualitative sample, laying the groundwork for further investigation of the phenomenon.

Knowledge about the investigated communities is not only necessary to understand and recognize specific linguistic patterns but also helps to differentiate societal norms from a potential self-classification of toxic language by users: Zhou et al., for example, note that mention of minority identities is often classified as toxic by current automated systems, while some of these words can express emotionality and closeness between the communicators (2021: 3143–3145). The same expression can be considered acceptable or toxic in different groups (Nobata et al. 2016: 146). Filters or lists are thus problematic if they don't distinguish the different assessments of toxicity in diverse communities (Mall et al. 2020: 2) and lead to false positives (discussed a.o. in Thompson et al. 2017). Multimodal toxicity is another challenge of Twitch streams, and especially image-based toxicity can stay undetected with text-based approaches that currently dominate toxicity research. To address these challenges of group-specific toxicity and multimodal data, the paper focuses on how toxicity arises across the different multimodal layers of the stream.

As a basis for the study, a stream sample was chosen that meets criteria favouring toxicity (massive stream size, male streamer, competitive online gameplay; Dreier, Pirker 2023) to identify various types of toxic actions and linguistic patterns. These patterns will be presented in examples, quantitative overviews and a tabular form to distinguish between targets, types, topics and linguistic/visual characteristics.

After giving a brief overview of current toxicity research in the context of online gaming and Twitch (2.) we will describe the data acquisition process and the chosen sample (3.). The discourse on the term 'toxic' in public discussions around streamers will be analysed (4.), setting the basis for the toxicity analysis (5.) that can help identify multimodal toxicity in streams. This analysis will focus on toxicity triggers (5.1), potential toxicity (5.2) and meta-discourse about toxicity (5.3).

2. Toxicity Research on Gaming and Twitch

In linguistic research, many automated approaches for toxicity detection have been developed either based on public discourse, with a focus on chat communication (Stoop et al. 2019 provide an early overview) or on communities that self-define as practising hate speech (Saleem et al. 2016). Piot et al. (2024), for example, combine over 1.2 million social media posts from prior studies to create the database MetaHate to further joint efforts in toxicity research. Though these quantitative approaches are constantly evolving and are already being implemented for automated toxicity detection on platforms, they also come with negatives like the aforementioned false flagging of minorities (Zhou et al. 2021), undetectable ambiguous meanings (e.g. Stoop et al. 2019: 20) and a lack of context in annotations that can result in false detection or overlooked toxicity. Sheth et al. (2022) therefore emphasise the value of qualitative analysis that can be combined with quantitative approaches.

Both Twitch and Online Gaming seem to be abundant sources for material in toxicity research due to the competitiveness of online gaming (Blackburn, Kwak 2014) and factors such as the online disinhibition effect (Beres et al. 2021). A common approach to research in these areas is the analysis of chat logs or automated log data in gaming (e.g. Martens et al. 2015; Thompson et al. 2017; Stoop et al. 2019). In a meta-study of 54 research papers on toxicity in online gaming, Anker Nexø (2024) identifies a current lack of multimodal approaches, an insight parallel to the review by Harpstead et al. (2019) of 46 studies of the platform Twitch. While there are already many 'technically supported' studies (which, for example, use metadata of the platform or chats) and 'technically agnostic' studies (such as surveys, for example, Cai, Wohn 2019), both Anker Nexø and Harpstead et al. describe how 'technically challenging' studies working with multimodal data are not yet as present but promise valuable insights.

Since toxicity is characterized by its property to spread in communication (Seering et al. 2017: 112), it is vital to see all toxic actions in their context, which, on Twitch, includes the broadcasted gameplay, interactions in the video/audio of the streamer and lastly the chat with its multimodal messages. Especially the emotes (images and GIFs used in chat by typing their names) differentiate the Twitch chat from many other platforms investigated in prior toxicity research, such as X (ElSherief et al. 2018) and Reddit (Mall et al. 2020). The reluctance for multimodal analysis of Twitch data on an extended scale can be explained by the difficulties in acquiring and transcribing the data (hence Harpstead et al.'s coining as "technically challenging studies" 2019: 6), which will be described in the following section.

3. Data acquisition and description of the sample

The aim of this paper is to describe different triggers (events or terms that can instigate or lead to toxicity) and multimodal ways to express toxicity. The objective for a data sample therefore was not to find a prototypical Twitch stream but to use a stream likely to be very toxic as a basis for the analysis. To find a suitable stream, the first step was to create a small sample (13 texts, 17,959 tokens) of meta discourse on toxic streamers by searching for the (potential) headline "Most toxic Twitch Streamers" (on the search engine Ecosia 04.07.25), checking for articles that mention or list specific streamers and their supposedly toxic behaviour.[1] The thirteen collected articles/blogs/forums were published between 2018 and 2024 (in the following called corpus B), and they list 90 different streamers, though only seven have had four or more mentions. Of these seven streamers (xQc, Ice Poseidon, Amouranth, Destiny, Forsen, JiDion, Pokimane), three were not active on Twitch (anymore) because they primarily produced content on YouTube or were banned from Twitch for violating community guidelines.

For all remaining four streamers, the allegations of toxic behaviour were checked in detail, and it became apparent that for the female streamers, their gender[2], and popularity on the platform might have contributed to the allegations. Thus, following the definition of toxicity in this paper (see 4.), neither female streamer met the criteria of interest and therefore did not seem promising for classifying different toxic actions. However, for the streamer xQc (mentioned in seven texts) and his community of 'Juicers' as well as Forsens 'Bajs' (four texts), the articles often describe both streamers and their communities as extremely toxic. According to Seering et al. (2017), streamers have a great impact as role models on their communities, so it is likely that toxicity will be a shared trait between streamers and their communities. It seems that "the group is defined by the speech and the speech comes to define the group" (Saleem et al. 2016: 2).

Dreier and Pirker (2023) identify gender, stream size and game genre as factors influencing toxicity levels in Twitch streams. In this regard, both male streamers with large audiences[3] are more likely to experience toxic incidents during streams. DotA 2 (Defense of the Ancients), a competitive online game considered very toxic in gaming communities and in research (Lee et al. 2021; Martens et al. 2015), was chosen for a stream sample since both streamers occasionally play the game. Based on a first multimodal impression of current streams from Forsen and xQc (e.g. use of swearwords, toxic messages in chat),

---

1 Articles that only focused on specific games and their communities were excluded as well as videos that were suggested by the search engine.

2 In a misogynist view of women as 'bad gamers' female live streamers allegedly use sexuality to attract audiences (and consequentially draw them away from serious streamers) while producing lower quality content. Community policing (for example flagging) or hostile raiding can therefore be used to exclude female streamers (see Adams et al. 2025: 8; Zolides 2023: 134).

3 At the time the data was collected (22.08.25), xQc was ranked 8th globally among Twitch streamers, and Forsen was ranked 251th on twitchtracker.com.

we decided on an xQc stream as a suitable sample, since he displayed more obvious toxicity on stream and seemed to have made toxicity a (potentially partially performative) part of his streaming style.[4] Because of the density of the material of a massive stream and the explorative character of this study, only one stream was analyzed, but the results will be discussed in future work in the context of nine more streams that are currently being collected.

Chosen as a sample for this study (corpus A) was a 140-minute stream of *DotA* gameplay from summer 2025,[5] which happened at the end of a 12-hour stream with various content. The late streaming time, as well as the two losses in the game, were assumed to contribute to increased toxic utterances and actions, which impacted the decision on this sample.

For the analysis, a transcript of the visual and auditive elements of the game, stream and chat was generated, utilizing Recktenwald's (2017) transcription scheme for Twitch. The chat messages, including time stamps from the Twitch stream, were copied to a txt-file and anonymized (using a Python script and following data protection guidelines). The chatters were assigned consistent identifiers (User+Number) to better understand shifts of behaviour over time; only the streamer and Bots (Schnozebot and Fossabot) kept their names. Moderators were marked as such since they can have a greater impact on both positive and negative community behaviour (Seering et al. 2017: 113). To simplify the annotation of the emotes, they were reduced to a text layer and were visibly marked through duplication of the emote's name (for example, LOLLOL). To better understand their content, all 890 emotes were additionally mapped in a CSV file to aid annotation and pre-sort potentially toxic emotes (see 5.2).

For the audio transcript, the automated transcript available on YouTube (of the reupload of the stream by xQc) was used as a basis, and the text was separated into timestamps (in seconds), a column for the streamers' (and other people's) utterances on stream, as well as the gameplay voices (characters, announcers, co-players). All swearwords that had been automatically censored

4 "It seems that being banned multiple times and cultivating a reputation as one of the most toxic streamers on Twitch is actually good for xQc, as he has almost 10 million followers on Twitch at the time of writing." (Dungca 2021)

5 Note on research ethics: Working with public data in the context of toxicity research can come with the stigmatisation of users, therefore the exact date of the chosen stream is not shared in this paper in order to not expose any viewers to potential negative consequences. For a further discussion on this topic see Legewie, Nassauer (2018).

by YouTube were uncensored, and mistakes of the transcript (like misspelling of in-game names, game-specific vocabulary or overlapping speech) were corrected manually following modified orthography transcript principles (eye-dialect, e.g. Edwards 1993: 11–20). A Python script was used to align timestamps and lines of text in a CSV file. In the next step, the titles of music played on the stream were manually added (using the app Shazam to identify unknown songs) as were visual descriptions of both the gameplay and the streamer. This resulted in a dense transcript of all three layers (game, stream, chat) as illustrated with a small sample in Table 1.

Tab. 1: Excerpt of the combined stream transcript.

Time	**Game**		**Stream**			**Chat**
	Visual	**Audio**	**Visual**	**Audio**		
				Voice	**Music**	
10:20:09	Wraith King (XQC) kills Ezalor (Enemy)				Culture Beat: Mr. Vain	User537: EZEZ, User463: LOLLOL, User23: EZEZ
[...]						
10:20:11				fucking loser	CB: MV	User238: wow, User538: EZEZ, User28: AYYYY, User66: EZEZ NICE, User82: nice
[...]						
10:20:13		WK: It was kind of you to grow that skeleton to maturity for me.			CB: MV	User142: AlienPlsAlienPls EDMEDM AlienPlsAlienPls EDMEDM AlienPlsAlienPls EDMEDM AlienPlsAlienPls EDMEDM AlienPlsAlienPls EDMEDM, User541: PogPog, User71: EZEZ, User89: LoserLoser, User98: EZEZ

Assigning every second of the sample one row in the CSV sheet (6,057 lines), the transcript contains the artist and title of identified music in the background, 18,200 chat messages (79,129 tokens), 855 lines of audio commentary from the streamer (5,916 tokens), as well as 411 descriptions of game visuals (3,279 tokens) and 389 of stream visuals (1,936 tokens).

4. Definition of the Term Toxic – Meta discourse

Toxicity can include all "actions of a user or users that create a negative atmosphere for the other users of social media" (Mall et al. 2020: 1). While this broad definition contains different multimodal forms of toxicity, it is also inevitably tied to a very subjective assessment, because the negative atmosphere must either be defined by the victims of a toxic action and thereby participants of a specific situation or through societal standards. The first viewpoint doesn't allow for judgement by external viewers, but societal standards, on the other hand, might not be appropriate or applicable to every specific community. Before annotating toxicity, therefore, insights from research as well as the public meta discourse on toxicity were evaluated, tying up to the three layers of the stream (game, streamer and chat):

Anker Nexø (2024) presents a list of 60 toxic actions discussed in analyzed research papers. These toxic actions range from unspecified toxicity, negative attitudes, to verbal actions (a.o. blaming, commands, complaints, cursing, flaming, hate speech, insults, leaking of information, malicious jokes, profanity, threats, name-calling or trash-talking), game specific behaviour (a.o. tea bagging, cheating, camping, intentional feeding, AFK-ing) or actions around the gameplay (a.o. lobby dodging, flagging, playing off-meta or smurfing). While the overview by Anker Nexø helps identify toxic game actions, it does not attempt to bring order to the often overlapping and synonymous descriptions of toxicity. Additionally, the streaming setting of the analyzed video brings another layer of potential toxic (non-verbal) actions with it that were, according to Nexø, not as present in prior research (2024: 7). Many studies on toxicity have a predefined concept of what is considered (and therefore annotated) as toxic instead of a data-driven approach that is followed here. For a less subjective classification of streaming-specific toxic behaviour, two sets of discursive data were analyzed to comprise a comparable list of typical toxic actions on Twitch. These negative actions were annotated using the software MAXQDA, following principles of the grounded methodology in an open annotation process.

While Twitch already implies certain undesirable actions in the Community Guidelines, acceptable or forbidden behaviour is often further specified by individual channel rules. In a preceding study[6] a corpus (corpus C) of over 50 channel rules with 3,517 tokens was annotated to find commonalities in explicitly undesired behaviour. The 490 annotations led to 44 common codes in eight primary and 36 subcodes. While many channels often use positive guidelines reminding of mindfulness/kindness, respect for streamers, viewers and moderators or having fun, toxicity and rudeness in general were often mentioned as prohibited actions. This includes specific practices like posting of (potentially dangerous) links, (self) promoting, mentioning of other streamers, doxxing, trauma dumping[7] or chat actions like spamming or the use of all-caps. Racism and sexism were usually general categories that were forbidden in any form; especially female streamers pointed out (sexual or any) comments on their bodies as reasons for bans. For practices more exclusively relevant to live-streaming, some streamers also explicitly prohibited 'backseating' or mentioning the stream's viewer count as personal preferences.

The discourse on toxic streamers (corpus B, described in section 3 containing 13 texts and 17,959 tokens) was also analyzed in MaxQDA by annotating and classifying all mentioned actions that could be considered toxic. 480 annotations distinguished between the topics of counter-actions against toxicity (5), results of toxicity (152) and toxic actions on Twitch in general (163), specifically by streamers (143) or viewers (17). Toxicity was described mostly in regard to discrimination or offensive speech, but there were also extreme examples discussing cases of pedophilia, suicide threats, sexual harassment or bullying. Often mentioned concerning streamers' behaviour on camera were animal abuse, aggressive actions, substance abuse, sexually explicit content, shaming/lashing out at viewers, but also the stirring of drama (to gain views) or hostile raiding. Especially actions related to finances, such as Ponzi schemes, promotion of cryptocurrencies, pressuring people into donating to the streamer

6 Presented by Bodden at GAL 2024 with the title "'No toxicity in chat' – Net-Etiquette im third space Twitch". The analyzed stream categories on Twitch were IRL, Minecraft, Fortnite, Music and Just Chatting. Especially Fortnite as the only clear competitive online gaming had the least number of guidelines in the checked accounts.

7 Meaning the (over)sharing of personal problems (e.g. relating to mental health) that brings a negative atmosphere into streams.

https://doi.org/10.14361/9783839477533-041 https://www.inlibra.com/de/page/agb - Open Access -

or gambling, were often mentioned as toxic practices that led to the banning of streamers from Twitch. Other toxic practices were attributed exclusively to viewers (e.g. swatting streamers, simping, spamming or making inappropriate requests), and it was discussed that toxic streamers attract dangerous communities that can show more extreme behaviour. A few of the analyzed articles discussed ways to prevent or counter toxicity, such as disengaging or directly calling out toxic actions. Most of them also mentioned the potential punishment for toxic actions in the form of banning, but also mental health problems, fines, the ending of careers or loss of monetary incentives were discussed.

The categories derived from the annotations of corpus B and C, alongside the reviewed literature, served as the basis for the following analysis of corpus A by sensitizing for different phenomena in online gaming and Twitch that can be understood as toxic.

5. Analysis of Triggers, Potential Toxicity and Meta Discourse

While the annotation was guided by the categories in the analyzed literature, guidelines and meta discourse, again an open coding process was followed where labels were newly generated, adjusted or merged over time, but followed the three guiding categories *Triggers* (5.1), *Potential Toxicity* (5.2) (here also differentiating between the layers game, stream and chat) and *Meta Discourse* (5.3). Additionally, further information, such as explanations of specific terms or practices, was added in memos (virtual notepads attached to a specific section of the transcript).

To better understand where toxic actions occur across the layers of the stream, the columns of the joined transcript were separated again into distinct documents that included the timestamp and text of each section, which already clearly indicate the density of potential toxicity (8,641 annotations, see 5.2) relative to tokens in each layer.

Tab. 2: Annotations and Potential Toxicity in the different layers of the transcript.

	Game		Stream		Chat	Total
	Visual	Audio	Visual	Audio (Voice only)		
Tokens	3,279	4,347	1,936	5,916	79,129	94,607
Total Annotations	249	226	4	382	8,695	9,556
Potential Toxicity	66	51	4	348	8,172	8,641

As shown in Table 2 in the game layer, most annotations were trigger events (5.1), and there were only a few toxic occurrences (usually in the form of chat messages by the co-players or taunts). The visual layer of the stream barely has any annotations, while the voice/audio part of the stream shows many annotations of toxic speech. The chat has both the highest number of tokens and with over 8 thousand annotations of toxicity around every 5th message contained toxicity.

The transcribed music cannot be discussed in detail in this paper and therefore does not appear in the overview. However, it seems fruitful to include this in a transcript because the communication often picks up the background music.[8] The levels of toxicity by the streamer and chat seem to loosely coincide with the tone/style of music in the background (e.g. more cursing with rap music than pop), which could be just a coincidence of the small sample and therefore has to be analyzed in detail in future work.

8 For example, someone insults the players as "99 NOOBS" (Chat, Pos. 838) tying up to the song '99 Luftballons' in the background. Especially the song 'Karma is my boyfriend' (or as someone rephrases, "Karma is losing Dota" Chat, Pos. 1072) by Taylor Swift sparks controversy in chat: When the streamer sings along parts of the lyrics, he is repeatedly labelled as a soyboy (a man who is showing more feminine traits than certain concepts of masculinity) and ridiculed.

5.1 Toxicity Triggers

Almerekhi et al. (2019: 291) describe Toxicity Triggers "in the context of online discussions as comments that incur direct toxic responses. [...] [T]hese triggers of toxicity may differ by the community and by topic due to different norms and uses of language". With a corpus of more than 104 million Reddit comments they identify that terms like "muslim, israel, kill, europe, support, tax, vote, law, president, woman" (ibid.: 292; also see TOXTRIG by Zhou et al. 2021: 3144) can lead to a toxic shift in comments. While political and religious topics are almost non-existent in the investigated sample,[9] other triggers, like for example 'president', were mentioned and triggered toxic reactions, but not in relation to real-world politics.[10] Other terms like 'kill', however, are very present in the gaming discourse of a MOBA (Multiplayer Online Battle Arena) but are not as likely to spark toxic reactions. A lexicon-based approach tied to these lists, therefore, does not seem promising.

Since research on toxicity in multiplayer online games agrees that it is usually in-game actions that form the basis for toxic behaviour, these were the focal points of the analysis. 374 potentially triggering game events were annotated across the game's layers (visual: 183; audio: 177) and stream (audio: 14). It is assumed that all events that are very negative in the game can lead to frustration in players, which increases the chance of toxic behaviour. Certain events (see table 2) are communicated in some form of visual or auditive ping or announcements, which were transcribed and annotated. This can help see which elements of the coded actions actually have an impact, or how, for example, the combination of a series of low-impact triggers can add up to a response that can be classified as potential toxicity (in 5.2). The impact of these events on the negative mood of the streamer can also be seen in the annotations of trigger events (here in the form of meta discourse).

9 Occasionally some viewers attempt to start a political discussion, but they seem to be lost in the masses of chat messages and usually do not serve as toxicity triggers "Is this the appropriate venue to share my Strongly held views on Israel?" (Chat, Pos. 2522). As Sjöblom & Hamari (2016) found out in their study about motivations for Twitch users, many people use streams for relaxation and don't want to encounter stressful topics. Consequentially, many streamers in the guideline corpus (C) explicitly forbid politics as a topic on their stream.

10 The streamer makes the proposition that he as a president would punish players that use taunts by turning of their utilities.

Tab. 3: Toxicity Triggers in the 3 layers of the stream.

Game visual	**Game Audio**	**Stream Audio (meta discourse)**
Lobby / Preparation Seeing a loosing streak Seeing bad matchmaking rank Queue resets because other player decline Seeing score board & match statistics Gold being lost by team for failure to pick a hero Chosen hero gets banned **In-game** Team members disconnect (& reconnect at a bad time) Player gets disconnected Mocking taunts by co-players Death of player / team members Courier being killed Towers being lost Enemies archive milestones Enemies' spells prevent player's actions Chat messages by co-players Announcements of enemies' success (double kill, triple kill, monster kill etc. by enemy team)	**Lobby / Preparation** Sound of gold being lost Countdowns for team actions (5 seconds remaining) Sound of hero being banned **In-game** Voiceline of character being under attack Announcements of enemies' success (double kill, triple kill, monster kill, killing Roshan) Announcements of team's losses (tower being attacked and falling etc.) Announcement of enemies' team win Voice taunts by co-players	**Lobby / Preparation** "this day I had a really bad loss streak" (Stream (Audio), Pos. 43)[11] "three in a row and I can't I I can't win" (Stream (Audio), Pos. 77) "oh my god we double picked him" (Stream (Audio), Pos. 240) **In-game** "is this guy gonna connect or what" (Stream (Audio), Pos. 475) "oh he reconnected (grunt) he reconnected at the last second that's absolute diabolical" (Stream (Audio), Pos. 589–594) "I'm lagging" (Stream (Audio), Pos. 3291)

It is assumed that all toxic actions can also function as toxicity triggers for further toxic behaviour tied to the broken window hypothesis (Cheng et al. 2017: 1219): especially when toxic actions go uncommented, this can be seen as an invitation for further toxic behaviour.

11 All quotes from the transcript will contain the layer (e.g. Chat / Stream (Visual)) and line (Pos.) in the transcript, spelling mistakes will not be corrected.

5.2 Potential Toxicity

The description of "Potential Toxicity" is based on the rejection of a definite attribution of 'toxicity' as a normative term that does not do justice to the various communities and their linguistic norms. Since the goal of this paper is not to judge certain behaviours but to annotate anything that can fall into a wide classification as 'toxic', there will be a tendency to over-annotate phenomena that can be mildly toxic or not considered problematic at all. For example, backseating (e.g. unwanted tips for the streamer's gameplay), which is explicitly forbidden in many individual community guidelines, is a common practice in the analyzed stream. Likewise, the use of swearwords (or abbreviations like wtf) is normalized within gaming communities. All these occurrences will still be considered as potentially toxic for this annotation. Other community-specific practices that seem to reside in a grey zone of toxicity are bantering and trolling, which the following example illustrates: People in chat start pointing out a loss in MMR (matchmaking rank) by stating numbers like "-50", indicating the streamer is not skilled, which can be toxic. This turns into a creative game where more and more absurd objects that are being lost are named. The commentators, therefore, make fun of the streamer (for being a bad player) as well as the chatters (who constantly need to point that out):

> User1182: 100 PLATES PogPog, User1608: -100 Airplanes / User786: xqcMxqcM 50 stones (Chat, Pos. 3523–3524) / User601: -100 seashells (Chat, Pos. 3527) / User8: POGGERSPOGGERS 500 clams (Chat, Pos. 3553) / User786: xqcMxqcM 50 granules of sand (Chat, Pos. 3554) / User309: 100 BABY TURTLES OOOOOOOO (Chat, Pos. 3556) / User1422: -38261 chicken nuggets LULWLULW (Chat, Pos. 3569)

While the streamer is usually ignoring the chat altogether and also does not seem to be bothered by negative messages when he sees them, the same kind of fun at another person's expense could be highly toxic in a different context. Displaying vulnerability online, especially as a streamer, is not always possible, so we cannot be sure of the impact of bantering and mockery. Normalizing these actions could negatively influence interactions in different communities as well. To investigate this, future research needs to engage with Twitch actors to better understand their attitudes and interpretation of potentially toxic actions, but insights from one community cannot be generalized easily.

5.2.1 Game

In addition to tagging (and classifying) toxicity, the annotation also includes the targets of a toxic action. These can be *directed* at a specific person (streamer, his girlfriend, teammates, enemies, other streamers, chat) but also *generalized* (for example, using ableist language) (Waseem et al. 2017). For this classification, the context and communication channels play an important role in identifying who is being addressed, for example, through pronouns.

In the game layer, players can use in-game chat, which is visible to either the team or the whole group of players or they can use taunts in both chat and game actions. Even positive or neutral taunts like "That was unreal" (Game (Audio), Pos. 2104) were used ironically to comment on a bad play by the streamer making them potentially toxic. The voice chat in contrast, is only audible to team members. While most of the communication channels can be blocked if a player is considered toxic, certain game actions (like throwing an item called 'Cloak of Intelligence' in front of another character to imply someone is stupid), pinging or drawing on the mini map are additional possible communication channels. Toxic players still find a workaround to be toxic with the communicative means available, though the means of communication for extremely toxic players (flagged by former co-players) will be restricted.

For *DotA*, characters' voice lines appear to be randomized. These voice lines usually are reactions to triggers and are often harmless bragging or taunting, but can sometimes contain curse words, like the character Wraith King saying, "Death is my bitch" (Game (Audio), Pos. 481), which is often picked up in the chat. Some voice lines also directly target opponents and can be classified as toxic: "You won't win the war by dying for your Ancient. You'll win by making the other dumb, unfed bastard die for theirs!" (Game (Audio), Pos. 3346)

5.2.2 Stream

The streamer has the most channels for communicating and the highest control of the communication situation, which increases the possible ways to be toxic. He can directly be toxic towards his teammates, enemies, viewers and people who are physically in his stream, in addition to talking about people in a toxic fashion. With the shared view of the game and the video of the streamer, other toxic non-verbal behaviour is also visible, for example, when he is baiting his teammates into a fight by asking for help and teleporting away, commenting on his actions in voice chat with "peace out" (Stream (Audio), Pos. 2350). In the stream, toxic utterances are usually directed at the teammates (with the primary topics skill and intelligence), or the game itself ("this is so goddamn fucking boring dude" (Stream (Audio), Pos. 5206)), and a lot of swear words are used (362 occurrences of swearwords without targeted toxicity in all layers).

The negative mentality towards team members is expressed, for example, when he comments "I don't give a fuck if my team dies" (Stream (Audio), Pos. 2372) or directly says in voice chat "you guys are awful by the way […] you're awful" (Stream (Audio), Pos. 2929–2931). As can be seen later on in the listing in Fig. 1, there are far more attacks on team members than enemies, which aligns with the findings by Martens et al. (2015: 5) that there is more toxicity towards team members, especially in losing games.

As we progress through the video, the frustration of the streamer becomes apparent in the visual layer, through fidgeting behaviour (resulting in the viewers spamming emotes like 'ADHD' displaying the streamer) or in aggressive gestures that sometimes emphasize toxic speech. Here, pitch, intensity of speech and prosodic markers of irony are impactful to classify toxicity and were partially noted in the transcript but would require further investigation. In terms of the visual layers of the stream (that barely received any annotations), especially the behaviour of the streamer AikoBliss, xQc's girlfriend, who is partially on camera, is interesting: One of the most viewed clips[12] on Twitch of the stream shows her in the background, exaggeratingly imitating his angry gesticulations to mock him while he complains about his co-players. Something similar happens in chat after xQc repeatedly urges his co-player (Shadow Shaman) to buy back[13] into the game to help the team (Stream (Audio), Pos. 2871) with higher and higher pitched voice and intensity that seems performative:

> Shaman buy back / Shaman buy back / Shaman buy back please man / Shaman please man buy back man / Shaman please buy back man / Shaman hehe Shaman buy back please man please man / Shaman buy back please man / Shaman please buy back man (Stream (Audio), Pos. 2871–2893)

It is assumed that toxic conduct by a streamer is not necessarily their authentic behaviour but can also be a form of performative toxicity for an audience.[14] The

12 If the clip function is enabled by streamers, any viewer can clip short sequences of a stream as highlights that can be shown on the streamers page but are also often circulated on other social media.

13 When characters die, players can pay in-game gold to make them respawn faster.

14 Anker Nexø points out that while Twitch streams give access to important audio-visual data that accompanies gameplay it also "risks biased data with incomplete, edited, se-

chat picks up this 'performance' by creating an echo that either repeats his pleas for support or distorts the utterance in a mocking fashion, as AikoBliss did with her gestures:

> User1303: shamaaan, User1150: BABAK, User697: LOLO, User357: DENTGE, User507: SHAMANBUYBACKMAN / User326: LOLLOL, […] User66: sha man, User82: omEomE, User8: wtf are u sayin LULWLULW […] User1205: BAKBAK MAN, User686: sher man / User246: LOLLOL, User697: hehe, User1252: sha man, User150: OMEGALULOMEGALUL, […] User813: PLEASEPLEASE, User768: LOLLOL […] User1194: wtf are you saying LOLLOL, User199: hehe, […] User1390: LOLLOL, User786: OMEGALULOMEGALUL, User1142: hehe, User571: DentgeDentge, User865: LOLO bebak, User28: LULLUL / User758: SHA MaNMaN, User11: LOLLOL hehe, […] User288: hehe / User357: LOLLOL, User663: LOLLOL, User304: sha MaNMaN, User1362: BAHBEHMEN, User45: LOLLOL / User813: hehe, User1391: SHA MAN, User49: LOLO, User659: Shamman, User276: he cant do anything without u, User1303: shemaaan (Chat, Pos. 2884–2891)

While the mocking in both instances could be potentially toxic, it can be assumed that this is bantering in the specific stream and community. The distortion effect of this bantering can however sometimes show ableist characteristics when impacted speech is associated with a lack of intelligence. This is condensed in the following example from chat where the insults "OkayegOkayeg beg reerrd playa" (Chat, Pos. 93) (directed at xQc) stands for 'big retard player' but also mimics an impairment in the spelling that is combined with an ableist emote.

5.2.3 Chat

With over 8,172 annotations of Potential Toxicity in the 18,200 chat messages by 2,506 different users, the stream's chat is the most important space for toxic behaviours. While both in the game and stream, skill and intelligence are predominantly the topics of insults, the themes become more diverse in the chat layer. Here, there are only two potential modalities available to express toxicity: written chat or emotes that can appear both as GIFs and images.

---

lected, or over-accentuation of behaviours. Hence, some video data may be produced with an audience in mind, which risks threatening the naturalistic emergence of behaviours and interactions displayed." (2024: 9)

As discussed before, emotes are a specific feature of Twitch compared to many other chat-based social media networks. Emotes can derive from three different types of sources: there are global Twitch emotes (like 'Kappa' or 'PogChamp'), streamer-specific emotes that are usually an incentive to subscribe to a channel by paying a monthly fee and lastly third-party extensions (for example BetterTTV, FrankerFazeZ or 7TV).[15] In the analyzed section of the stream, there were 890 different emotes used, which were manually annotated in a mapping document. By looking up origins, meanings and context of the emotes and their usage in chat, they were classified as potentially toxic when applicable. This was used as a basis to automatically annotate thousands of emotes throughout the document. For example, the emote 'GYATT' shows the streamer with an exaggerated expression, looking at something. While gyatt as a term can be an expression of being impressed in contemporary youth language, its origin refers to admiring female behinds, so the term is intertwined with highly misogynistic concepts of male gazing. This reading of the emote can be confirmed in the sample where 'GYATT' is posted mostly in a situation when a player's profile in the game (that is customized) shows a selection of the hovering cleavages of several female DotA characters.

Some of the emotes have multiple topics and targets that were annotated additionally. For example, the ableist emote 'Dentge' (a frog and a variation of the Pepe-meme with an indented head implying a physical and mental disability) is generally toxic by targeting disabled people, but simultaneously can be used to express, for example, a lack of skill in a co-player.

The emotes on Twitch are considered memes, since they are communicative elements that are based on a creative use of intertextuality (specifically meme patterns) and virally distributed by numerous users (Shifman 2014: 44). Just like classic meme templates, they can fulfill important functions of mimic and gestures in the written conversation in the chat since "even the snapshot depictions of embodied features—including facial expressions, gestures, (changes in) postures—however much reduced they may be compared to full face-to-face interaction, contribute much to the emergence of meaning, especially stance meanings" (Dancygier, Vandelanotte 2025: 13). Most of

15 The first type of emote can be used and be seen by everyone, community emotes are curated by the streamers and are only available to subscribers but can be seen by everyone and lastly the third party extensions can be generated by every user but they are not being displayed for anyone who does not use the extensions so they can appear as (sometimes incomprehensible) text.

the memes, therefore, show actual people, animals or cartoon characters to express some form of emotion.

But this also means that any emote (though it might seem toxic at first glance) can be used in very different ways depending on its context, which can be described as a "memetic grammar" (ibid.: 4). For example, the combination of the frog-emote 'okayeg' and a clock emoji can be read as 'retard time' insulting the players of the prior game actions, but 'okayeg' can also be used to convey doubt in another context. The comment "MOST INVOKERS 5Head5Head ThinkingThinking DentgeDentge Thinking2Thinking2" (Chat, Pos. 3315) uses four combined emotes to draw a head next to a thinking bubble showing the physically impaired 'Dentge' frog. The difficulty with the emote '5Head' on Twitch is that it can both stand for someone very smart or very stupid. The chat messages seem to use this ambiguity to express a message like 'Most players of the character Invoker think they are very smart, but they are actually very dumb (/disabled)'. The knowledge of the game, the community, and the context is, therefore, necessary to correctly estimate the intended meaning for this kind of message, which underscores the approach of classifying everything that could be toxic as potentially toxic. Emotes can also share positive connotations, imply irony or change the meaning of a verbal sentence, which was already considered for emojis in an automated classifying approach by Thompson et al. (2017: 151).

While it is easy to spot clearly racist, sexist or ableist emotes in the analyzed material, there are two problems for efforts to restrict this form of toxicity: First of all, many users might not know the toxic background of a specific emote (for example emotes featuring topics like 'soy boys' or 'clown world' relating to AltRight movements[16]) or they do not care about their origin and hurtful connotations. This becomes very clear when looking at the list of hosted emotes by the streamer.[17] But also apart from emotes, the chat can contain visual elements comprised of graphemes by using ASCII art. This was enabled by the

16 See Liemann (2023) for a deeper discussion of alt-right memes relating to gender.

17 For an overview see: https://www.xqcupdates.com/emotes. It is noteworthy that many of the curated emotes show the streamer himself tagged with derogatory terms like Loser, Jackass, Cucklord or RetardJam.

streamer through the automated Schnozebot, for example, when a user writes the command "!lewd" in chat, the bot draws an Ahegaho-face.[18]

Additionally to the visual elements in chat, of course, most toxicity was expressed verbally. For example, different kinds of curse words appear in chat unfiltered but often abbreviated. In addition to these slur words, derogatory terms (for example, the streamer is often called bub, good boy, lil gup, lil buddy) or dehumanizing metaphors (e.g. animal, doggie, roach, monkey or bot) are used to either attack people in the game or the streamer. On the level of sentence structures, toxicity can also be expressed, for example, by the usage of repetition and distortion, which was already visible in the 'buy-back-Shaman' example.

Complex forms of intertextuality are another resource that is used to create new forms of creative insults. When the streamer, for example, is not playing the game according to the chat's standards, there are repeated comments about a POV (point of view, a term from digital storytelling culture), meaning he only shows the game from a non-player perspective or commenting on him only walking instead of fighting, indicating his lack of skill:

> User1358: Cameraman POV (Chat, Pos. 5120) / User1925: sakura pov LOLLOL (Chat, Pos. 5307) / User1932: COURIOUR POV LOLLOL (Chat, Pos. 5312) / User2074: you out there doing cardio prePffttprePfftt (Chat, Pos. 5321) / User1861: walking sim (Chat, Pos. 5125) / User2015: walking ward (Chat, Pos. 5120)

Many chat messages in massive Twitch streams are highly intertextual and might not seem to make sense at first glance. For the disambiguation of both emotes and written text in chat, Ford et al.'s conceptions of 'crowdspeak' and the three practices that create coherence (2017: 863) in streams can therefore be helpful. They describe that "massive chats are filled with non-sequiturs, verbatim repetition, variations on prior messages, blocks of text copied from elsewhere, and tiny messages that include only a few words or emotes" (ibid.: 859). 'Shorthanding' (the contraction of text into a smaller space[19]) and 'bricolage'

18 A special term for (usually female) faces expressing pleasure in Japanese animated porn. Ahegaho-faces are explicitly forbidden in Twitch emotes which, in this case, is circumvented by the use of the bot.

19 For example, the message "ur cm afk om" (Chat, Pos. 760) stands for 'your Crystal Maiden (a character in *DotA*) is away from their keyboard' and om is likely standing either for OMG (oh my god) or 'OmE' (short for Omegalul) used as an intensifier to ex-

(the recombining of elements from a small repertoire[20]) can be seen, for example, in the described practice of distortion and are guiding principles for both text messages and meme culture. But especially the concept of 'voice taking' can be helpful for toxicity annotation: 'Voice taking' describes the adoption of shared viewpoints and mannerisms by people in chat, which means that not as many new thoughts are being presented in chat, but they are just repeated in variations. Massive chats can therefore "maintain[] coherence by reducing the total volume of meaningful content participants produce and process" (Ford et al. 2017: 860). For toxicity annotation, this means that messages close to one another are likely to relate to the same subject and might therefore have comparable levels of toxicity.

For example, the following chat messages all contain the same meaning (ridiculing xQc's hair, appearing in about 320 comments in total) using emotes, sarcasm, comparisons or metaphors to convey a toxic message:

> User578: LOLLOL HAIR (Chat, Pos. 5831) / User1313: xqcTreexqcTree (Chat, Pos. 5832) / User786: bird hair (Chat, Pos. 5833) / User8: NAHHNAHH WTF IS THAT HAIR (Chat, Pos. 5834) / User1470: holy hair OMEGALULOMEGALUL (Chat, Pos. 5835) / User2261: forsenLaughingAtYouforsenLaughingAtYou HAIR (Chat, Pos. 5835) / User278: LOOK AT THAT BANAN PEEL (Chat, Pos. 5837) / User1473: brother cut that shit (Chat, Pos. 5836) / User64: omEomE nice hair (Chat, Pos. 5838) / User2301: WutFaceWutFace look at him (Chat, Pos. 5841) / User2303: (palmtree emoji) ClapClap (Chat, Pos. 5844) / User2274: MOPHEAD ANDY LOLLOL (Chat, Pos. 5857) / User681: @xQc brother that hair makes you look like a mushroom (Chat, Pos. 5873)

5.2.4 Quantitative overview of potential toxicity in all layers

As visible in Fig 1. the streamer is the most popular target of toxicity since he is the center of attention for anything happening in chat. It is noteworthy that this is often the result of a sort of toxic echo: Whenever he brags, down-

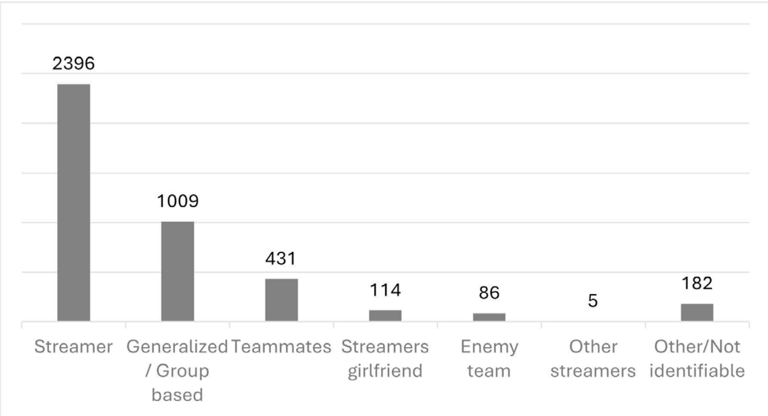
press excitement or Schadenfreude. While it could be a literal interpretation of the in-game action, the term afk is often used to say someone is so bad it is equal to them not playing at all. In another comment someone writes „alt+f4" (Chat, Pos. 2953) (a command on the pc to shut the window you are currently using) to say 'stop playing the game'.

20 An example for this would be the variations of LOL, LUL, OMEGALUL or omE (all expressing surprise or enjoyment) that can be found both in emotes or texts.

https://doi.org/10.14361/9783839477533-041 https://www.inlibra.com/de/page/agb - Open Access -

talks something, or attacks co-players, the chat will either join in on the attack or will react in a counterattack. There is also a high degree of group-based or generalized toxicity, for example, in hate speech regarding racism, sexism and ableism. The streamer's girlfriend AikoBliss is only partially visible during the stream but receives a lot of attention, both with positive and negative messages. While there are some comments that can be read as sexist, most of them request her to do something (for example, strike a pose with the emote 'PotFriend').

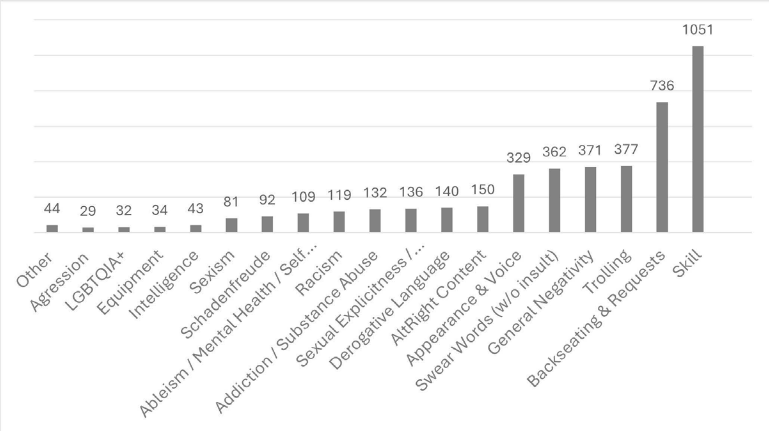
Fig. 1: Targets of potential toxicity



Some of the presented examples in this text might seem comparably harmless but there are also very toxic comments, for example, in the form of explicit sexual harassment, especially towards the male streamer, which often only become apparent through context. For example, when he says about the game "I am actually so solid I am going to bust everybody" (Stream (Audio), Pos. 1349), which already has sexual connotations, a few people in chat answer "me next" (Chat, Pos. 1361) or post more explicit content.

Through the open annotation process, a mix of categories was annotated that can provide a better insight into the different content and their quantity:

Fig. 2: Annotated themes of potential toxicity.



As visible in Fig. 2, most of the potential toxicity relates to the game by criticizing skill or backseating[21] but also trolling and swear words as mild forms of toxicity are very present. Under the term 'General Negativity', different types of toxic actions that can spread a negative mood (for example, complaining about the chosen game) were annotated, but also attacks that did not fit any of the other categories and seemed too minor to start a new section. Throughout the annotation process it became apparent that the categories have very different content because they described both types of actions, like trolling or backseating and topics such as racism. This led to annotations sometimes overlapping different categories, which affected the clarity. For further research with a qualitative approach, it therefore could make sense to differentiate not only between targets but also types and topics of a toxic action. Furthermore, a classification can also include a description of the linguistic or visual means that were used to convey the toxicity, as discussed in the examples before:

21 For this, only sentences with explicit imperatives were counted and polite suggestions or single words (like the name of an item) were excluded and would have resulted in a much higher number.

Tab. 4: Multimodal Annotation Grid

Target	Type	Topic	Linguistic / Visual Means
Generalized Directed unknown, streamer, chat, co-players (enemy & team), additional streamers, people on stream	spam, backseating, requests, aggressive behaviour, inappropriate language, (ad-hominem) attacks, critique, doxxing, Schadenfreude, down talking something	racism, sexism, ableism, LGBTQIA+, politics, religion, appearance, ageism, mental health, skill, equipment, intelligence	Linguistic swear words, derogative terms, raised voice, rhetorical questions, anaphor, 'easy language', sarcasm, verbal irony, homophony Both distortion, intertextuality, repetition, metaphors, ASCII art Visual mimic, gestures, movements, images, all caps

5.3 Meta Discourse

To better understand the perceptions and mechanisms of toxicity on Twitch, meta-discursive discussions related to toxicity were also annotated. This includes comments that call out toxic behaviour, counter speech, comments that instigate more toxic behaviour or reflections on moderation or community norms.

Some people in chat tried to counter toxicity by making it explicit, which could enable a discussion of appropriate behaviour. In the annotated material, four different types of meta-communication about the toxicity of the streamer were identified:

Tab. 5: Reflections on toxic behaviour in chat

Irony	Statement	Perspective of victim	Requests
LOLLOL toxic much (Chat, Pos. 2990) nothing better thn a bit of toxicity in dota (Chat, Pos. 5802) W TOXIC (Chat, Pos. 2696)	rude to your team WeirdChamp-WeirdChamp (Chat, Pos. 2044) D:D: so mean (Chat, Pos. 2841) ur soo toxic (Chat, Pos. 3782) minus behavior score (Chat, Pos. 5626)	True but poor him (crying cat emote) xqc (Chat, Pos. 5620) Buddy comes home from work, he finds a dota 2 match and instantly gets into a match with his favourite streamer. He picks support to support his streamer. Streamer blames him. Buddy becomes sad. Buddy no longer wants to play. (Chat, Pos. 1053)	be nice (Chat, Pos. 5508) BEAN ICE SAJSAJ (Chat, Pos. 2053) SAJSAJ apologize (Chat, Pos. 5533) D:D: dont be mean (Chat, Pos. 2848) @xQc Dont be rude mofo (Chat, Pos. 5524)

While one user asks xQc to watch his language specifically for a younger audience[22] sometimes his use of milder curse words is also commented upon since his audience expects a certain kind of language.[23] Especially the second example that takes the perspective of the victim of an attack in table 5 was posted repeatedly in chat and creates a narrative about his co-player suffering from the toxicity (Chat, Pos. 1053). While the streamer does not pay attention to the chat for most parts of the stream and will not likely be influenced by this attempt of counter speech, it can show people in chat that not everyone accepts toxic behaviour, which can have an impact on regulating the community.

But the streamer himself also partially reflects on the limits of what is sayable: While he queues for a game on a South American server he comments: "South American Dota by the way is– I can't I don't know if I can say what I want to say– like it's just it's just it's just special" (Stream (Audio), Pos. 175–180).

22 "WeirdCatWeirdCat stop swearing all the time my kids are watching" (Chat, Pos. 5207). The stream was not marked as +18 and therefore should not contain swearwords in accordance with Twitch's community guidelines.

23 One chat message points out that xQc's way of speaking sounds like the community of the competitive online game Valorant ("HE HAS THE VAL ACCENT LMFAOOLMFAOO" (Chat, Pos. 5613)), which is connected to the attacking of co-players (Çakır 2023).

While this sparks countless of racist comments towards South American players, it also opens a meta discourse where people urge him to clarify ("UNIQUE OR SPECIAL?" (Chat, Pos. 183), "wdym SusSus" (Chat, Pos. 186), "say it" (Chat, Pos. 185)) or give him the free pass to be clearly racist ("i won't tell anyone if you say it" (Chat, Pos. 182)). The chat is also flooded with the 'IFISPEAK' emote,[24] and xQc receives the advice to "tread carefully" (Chat, Pos. 209) due to the potential backlash. This is made more explicit by a comment about a fellow streamer who faced consequences after talking negatively about Brazilians: "bro u gonna get hated like Zombs for no reason (skull emoji)" (Chat, Pos. 203). It is not clear, however, if viewers actually disagree with potentially racist utterances or just fear for their streamer to get punished (e.g. with a shitstorm) for his behaviour.

This also becomes apparent when a user writes "XQC you should say 'OSUJDAYU' when he says such a things on stream" (Chat, Pos. 5429), which stands for the Russian word осуждаю and means something like 'I judge' and is commonly used as a defence for offensive language so streamers don't get banned for the behaviour of their community or co-players. This practice of claiming to judge someone for toxicity seems to be a ritualized defence mechanism without the necessity to believe in the statement.

There is an ongoing interaction in chat between users and moderators about accepted behaviour and forms of punishment, for example, when viewers request or question decisions for banning:

> User 266: @User1121 mods GTABGTAB (Chat, Pos. 1713) / User266: @User1022 MODSMODS BANDBAND BANDBAND (Chat, Pos. 1793) / User2398: @xQc MOD HIM (Chat, Pos. 5934) / User73: @User323 game complainer MODSMODS (Chat, Pos. 795) / User438: holy shit he said that and didn't get banned i said the word and got 5 year ban (Chat, Pos. 5964)

Especially the emotes 'GTAB' (get that ass banned), 'Band' (showing a band but using homophony) and 'MODS' (to inform moderators) are used in this context, but this practice is also seen as negative actions of a "snitch @User73 :(:(" (Chat, Pos. 814).

The collected material (though only discussed here with examples) shows there is an ongoing awareness and debate even in a seemingly toxic community

24 It features the football coach Jose Mourinho during a post-game interview. He was asked a controversial question and said, "If I speak I am in big trouble".

about what is sayable both in chat and for the streamer. Moderation actions like banning can help regulate behaviours in chat (Seering 2020) but are often not visible for the rest of the community. While the different types of counter speech towards the streamer don't seem to have a great impact on his behaviour they defend values of civilized speech. These actions can be impactful for the community, as challenging toxic practices makes user's disagreement visible and might create awareness in more viewers to question them. But as an article from the corpus B (Shrivastava 2022) reports, xQc has in the past banned users who call him toxic from his stream and labeled them 'Social Justice Warriors' (a term that ridicules people who stand up for more progressive causes), which is another form of regulating what is sayable in this community.

6. Conclusion

The article gave a first impression of the many challenges and the great potential to investigate toxicity in multimodal streams on Twitch. The paper uses insights from toxicity research both on Twitch and in gaming as a basis to expand it with a multimodal approach. For this, a methodology to transcribe Twitch streams was developed, and a suitable sample (corpus A) was found through discourse analysis. In order to be mindful of the different conceptions of what can be considered as toxic, research articles, Twitch communities (corpus C) and an excerpt of the public discourse (corpus B) were analysed. This led to the decision to annotate all cases of *potential* toxicity and to conceptualize toxicity tied to Wittgenstein as a *Sprachspiel* that can be a communicative practice.

In the analysis, the three focal points of Triggers, Potential Toxicity and Meta Discourse were discussed using examples from the corpus and a quantitative overview. A main distinction between existing research on triggers in the context of gaming is that multimodal data shows all actions that can elicit negative feelings in players, allowing the impact of smaller triggers to be studied in more detail.

For annotating Potential Toxicity, the analysis was broken down into the three layers of game, streamer and chat, but still reflected their interconnectedness. In the game layer, multiple actors were able to use voice chat, written chat or in-game actions like taunts to attack each other. It was mostly teammates being toxic to one another rather than enemy players, and the streamer was the only person who was directly toxic in voice chat.

For the stream layer, it was shown that both game actions and the verbal layer were used for toxicity, but also generalized toxicity, like the usage of swear words, was very common. For the visual layer, the video of the streamer illustrated the shift to a negative mood and aggressive gestures intensified toxic speech. The style of bantering in the interaction of the community also became apparent, for example, in the streamers' girlfriend mocking his complaints about co-players with exaggerated gestures.

Regarding the chat, emotes (as a form of meme culture) and 'crowd speak' were discussed to understand the interaction dynamics between viewers. Principles like repetition, distortion and voice taking helped to classify messages that can be highly toxic and would not be easy to detect with current approaches. The chat had diverse topics, but the primary focus of all toxic attacks was the streamer himself, who was criticized for his gameplay, appearance, taste in music or behaviour. Whenever he attacked co-players, it resulted in a form of echo that would partially target his co-players but mostly fire back at him, which could be interpreted as some form of punitive mechanism for toxic behaviour. This also becomes apparent in the analyzed metadiscourse, where the streamer either gets warned of consequences, there are attempts to push him to overstep by being explicitly racist or forms of counterspeech in chat.

Due to the sheer amount of data, many phenomena could not be discussed here but will be the object of future research, which will, for example, demonstrate the interconnectedness of the three layers of streams in a map visualisation (showing all trigger events and toxicity in the three layers according to the timeline). Another step is also to compare this manual annotation process and its results with automated approaches and to investigate the functions of the observed behaviours for community building.

Bibliography

Adams, Philippa R.; Sommers, Maria; Scholl, Ben; Lusoli, Alberto; Chun, Wendy H. K. (2025): The revenge of 'tribal' toxic masculinity: hate raids on Twitch. In: Journal of Visual Culture. 24, 1. URL: https://doi.org/10.1177/14704129251327916.

Almerekhi, Hind; Kwak, Haewoon; Salminen, Joni; Jansen, Bernard J. (2019): Detecting Toxicity Triggers in Online Discussions. In: Proceedings of the

30th ACM Conference on Hypertext and Social Media (HT'19). Hof, Germany. P. 291–292. URL: https://doi.org/10.1145/3342220.3344933.

Anker Nexø, Louise (2024): Toxic Behaviours in Esport: A Review of Data-Collection Methods Applied in Studying Toxic In-Gaming Behaviours. In: International Journal of Esports. 1, 1. URL: https://www.ijesports.org/article/127/html [accessed 03.02.26].

Beres, Nicole A.; Frommel, Julian; Reid, Elizabeth; Mandryk, Regan L.; Klarkowski, Madison (2021): Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. ACM New York, NY, USA. P. 1–15.

Blackburn, Jeremy; Kwak, Haewoon (2014): STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games. In: WWW '14: Proceedings of the 23rd international conference on World wide web. P. 877–887. URL: https://doi.org/10.1145/2566486.2567987.

Cai, Jie; Wohn, Donghee Y. (2019): What are Effective Strategies of Handling Harassment on Twitch? Users' Perspectives. In: Conference companion publication of the 2019 on computer supported cooperative work and social computing. P. 166–170.

Çakır, Gökhan (2023): What is the VALORANT accent? Explained. "Toxicity of our city". In: Dotesport.com. URL: https://dotesports.com/valorant/news/what-is-the-valorant-accent-explained [accessed 08.05.26].

Cheng, Justin; Bernstein, Michael; Danescu-Niculescu-Mizil, Christian; Leskovec, Jure (2017): "Anyone can become a troll: Causes of trolling behavior in online discussions". In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17. New York, NY, USA: Association for Computing Machinery. P. 1217–1230. URL: https://doi.org/10.1145/2998181.2998213.

Dancygier, Barbara; Vandelanotte, Lieven (2025): Embodiment and simulated interaction in online stance expression. In: Front. Psychol. 15, 1479825. URL: https://doi.org/10.3389/fpsyg.2024.1479825.

Dreier, Lukas; Pirker, Johanna (2023): Toxicity in Twitch Live Stream Chats: Towards Understanding the Impact of Gender, Size of Community and Game Genre. In: 2023 IEEE Conference on Games (CoG), 21.08.2023, 1. URL: http://dx.doi.org/10.1109/CoG57401.2023.10333159.

Dungca, Gio (2021): 5 Most Infamous Gaming Streamers of All Time In: Clutchpoints. URL: https://clutchpoints.com/gaming/5-most-infamous-gaming-streamers-of-all-time [16.05.26].

Diwanji, Vaibhav; Reed, Abigail; Ferchaud, Arienne; Seibert, Jonmichael; Weinbrecht, Victoria; Sellers, Nicholas (2020): Don't just watch, join in: Exploring information behavior and copresence on Twitch. In: Computers in Human Behavior. 105, 106221.

Edwards, Jane (1993): Principles and Contrasting Systems of Discourse Transcription. In: Talking Data. Transcription and Coding in Discourse Research. Ed. by Jane A. Edwards & Martin D. Lampert. Hillsdale: Lawrence Erlbaum Associates Inc. P. 3–31.

ElSherief, Mai; Nilizadeh, Shirin; Nguyen, Dana; Vigna, Giovanni; Belding, Elizabeth (2018): Peer to Peer Hate: Hate Speech Instigators and Their Targets. In: Proceedings of the ICWSM 2018. 12, 1. URL: https://arxiv.org/pdf/1804.04649v1 [accessed 25.04.26].

Ford, Colin; Gardner, Dan; Horgan, Leah Elaine; Liu, Calvin; Tsaasan, A. M.; Nardi, Bonnie; Rickman, Jordan (2017): Chat Speed OP PogChamp: Practices of Coherence in Massive Twitch Chat. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems – CHI EA '17. P. 858–871. URL: https://doi.org/10.1145/3027063.3052765.

Harpstead, Erik; Sebastian Rios, Juan; Seering, Joseph; Hammer, Jessica (2019): Toward a Twitch Research Toolkit: A Systematic Review of Approaches to Research on Game Streaming. In: Proceedings of the Annual Symposium on Computer- Human Interaction in Play. P. 111–119.

Kwak, Haewoon; Blackburn, Jeremy (2014): Linguistic analysis of toxic behavior in an online video game. In: Social Informatics. SocInfo 2014. Ed. Luca Aiello, Daniel McFarland. Springer, Cham. URL: https://doi.org/10.1007/978-3-319-15168-7_26.

Legewie, Nicolas; Nassauer, Anne (2018): YouTube, Google, Facebook: 21st Century Online Video Research and Research Ethics. In: Forum Qualitative Sozialforschung / Forum: Qualitative Social Research. 19, 3 URL: https://doi.org/10.17169/FQS-19.3.3130.

Liemann, Christina (2023): Hysterische Feministinnen, Soy Boys und SJWs. Geschlecht als multimodale Argumentationsstrategie in rechts-alternativen YouTube-Videos. In: Aptum. Zeitschrift für Sprachkritik und Sprachkultur. 19, 1. P. 2–17.

Lee, Matthew; Johnson, Daniel; Tanjitpiyanond, Porntida; Louis, Winnifred R. (2021): It's habit, not toxicity, driving hours spent in DOTA 2. In: Entertainment Computing, 100472. URL: https://doi.org/10.1016/j.entcom.2021.100472.

Mall, Raghvendra; Nagpal, Mridul; Salminen, Joni; Almerekhi, Hind; Jung, Soon-Gyo; Jansen, Bernard J. (2020): Four Types of Toxic People: Characterizing Online Users' Toxicity over Time. In: Association for Computing Machinery. URL: https://doi.org/10.1145/3419249.3420142.

Martens, Marcus; Shen, Siqi; Iosup, Alexandru; Kuipers, Fernando (2015): Toxicity detection in multiplayer online games. In: 2015 International Workshop on Network and Systems Support for Games (NetGames). P. 1–6. URL: https://doi.org/10.1109/NetGames.2015.7382991.

Nobata, Chikashi; Tetreault, Joel; Thomas, Achint; Mehdad, Yashar; Chang, Yi (2016): Abusive language detection in online user content. In: Proceedings of the 25th International Conference on World Wide Web, ser. WWW '16. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. P. 145–153. URL: https://doi.org/10.1145/2872427.2883062.

Piot, Paloma; Martín-Rodilla, Patricia; Parapar, Javier (2024): MetaHate: A Dataset for Unifying Efforts on Hate Speech Detection. In: Proceedings of the International AAAI Conference on Web and Social Media. 18, 1. P. 2025–2039. URL: https://doi.org/10.1609/icwsm.v18i1.31445.

Recktenwald, Daniel (2017): Toward a Transcription and Analysis of Live Streaming on Twitch. In: Journal of Pragmatics. 115. P. 68–81.

Saleem, Haji M.; Dillon, Kelly P.; Benesch, Susan; Ruths, Derek (2016): A web of hate: Tackling hateful speech in online social spaces. Proceedings of Text Analytics for Cybersecurity and Online Safety (TA-COS), LREC Portoroz. P. 1–9.

Seering, Joseph (2020): Reconsidering community self-moderation: The role of research in supporting community-based models for online content moderation. In: Proceedings of the ACM on Human-Computer Interaction. 4, 107. https://doi.org/10.1145/3415178.

Seering, Joseph; Kraut, Robert; Dabbish, Laura (2017): Shaping pro and antisocial behavior on twitch through moderation and example-setting. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17. New York, NY, USA: Association for Computing Machinery. P. 111–125. URL: https://doi.org/10.1145/2998181.2998277.

Shrivastava, Aarnesh (2022): Twitch streamers who absolutely lost it in front of their viewers during a livestream. In: sportskeeda. URL: https://www.sportskeeda.com/esports/5-twitch-streamers-absolutely-lost-viewers-live-stream [16.05.26].

Sjöblom, Max; Hamari, Juho (2016): Why do people watch others play video games? An empirical study on the motivations of twitch users. In: Computers in Human Behavior. 75. P. 985–996. URL: http://dx.doi.org/10.2139/ssrn.2779543.

Sheth, Amit; Shalin, Valerie, L.; Kursuncu, Ugur (2022): "Defining and detecting toxicity on social media: context and knowledge are key". In: Neurocomputing. 490. P. 312–318. URL: https://doi.org/10.1016/j.neucom.2021.11.095.

Shifman, Limor (2014): Memes in Digital Culture. Cambridge, Massachusetts: The MIT Press.

Stoop, Wessel; Kunneman, Florian; van den Bosch, Antal; Miller, Ben (2019): Detecting harassment in real-time as conversations develop. In: Proceedings of the Third Workshop on Abusive Language. P. 19–24. URL: https://doi.org/10.18653/v1/W19-3503.

Thompson, Joseph J.; Leung, Betty H. M.; Blair, Mark R.; Taboada, Maite (2017): Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model. In: Knowledge-Based Systems. 137. P. 149–162. URL: https://doi.org/10.1016/j.knosys.2017.09.022.

Twitchtracker: Twitch Statistics & Charts. URL: https://twitchtracker.com/statistics [15.05.26].

Waseem, Zeerak; Davidson, Thomas, Warmsley, Dana; Weber, Ingmar (2017): Understanding abuse: A typology of abusive language detection subtasks. In: Proceedings of the First Workshop on Abusive Language Online, Vancouver. P. 78–84.

Wittgenstein, Ludwig (1967): Philosophische Untersuchungen. Frankfurt a.M. 1967.

Zhou, Xuhui; Sap, Maarten; Swayamdipta, Swabha; Choi, Yejin; Smith, Noah (2021): Challenges in automated debiasing for toxic language detection. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. P. 3143–3155. URL: https://doi.org/10.18653/v1/2021.eacl-main.274.

Zolides, Andrew (2023): Toxic community policing: Weaponizing moderation tools on Twitch. In: Real Life in Real Time: Live Streaming Culture. Ed. by Johanna Brewer, Bo Ruberg, Amanda L. L. Cullen, Christopher J. Persaud. Cambridge, MA: MIT Press. P. 131–144. URL: https://doi.org/10.7551/mitpress/14526.003.0013.