Lost in the Crowd: How Group Size and Content Moderation Shape User Engagement in Live Streaming

Keran Zhao,[a] Yili Hong,[b,*] Tengteng Ma,[c] Yingda Lu,[d] Yuheng Hu[e]

[a] Department of Supply Chain and Information Systems, Smeal College of Business, Pennsylvania State University, State College, Pennsylvania 16802; [b] Department of Business Technology, Miami Herbert Business School, University of Miami, Coral Gables, Florida 33146; [c] School of Information Systems and Management, University of South Florida, Tampa, Florida 33620; [d] Department of Information and Decision Sciences, College of Business, University of Illinois Chicago, Chicago, Illinois 60607; [e] Department of Accounting and Management Information Systems, Fisher College of Business, Ohio State University, Columbus, Ohio 43210
*Corresponding author
Contact: krzhao@psu.edu, https://orcid.org/0000-0002-0142-4572 (KZ); khong@miami.edu, https://orcid.org/0000-0002-0577-7877 (YH); tengtengma@usf.edu, https://orcid.org/0000-0003-1822-3168 (TM); yingdalu@uic.edu, https://orcid.org/0000-0001-9482-6759 (YL); hu.3331@osu.edu, https://orcid.org/0000-0003-3665-1238 (YH)

Abstract. Despite the focus of the digital content literature on asynchronous platforms (e.g., Reddit, Wikipedia, Yelp), synchronous content platforms like live streaming (e.g., Twitch, Youtube Live) have become increasingly popular for enabling real-time user engagement at scale. These platforms involve engaging a sizable user base, facilitating their interactions within a contemporaneous environment. As group size increases, real-time interactivity scales rapidly, making the engagement interface fast-paced and erratic. Consequently, live-streaming channels often use moderators to address the challenge. Although the existing literature finds that the audience's group size has a positive effect on user engagement on asynchronous platforms, how group size affects synchronous interactions, particularly with the presence of bot and human moderators, is unclear. In this work, we leverage exogenous increases in live streaming viewers (from the Raid function in Twitch) to empirically examine the impact of group size on commenters' engagement in real time. Furthermore, we delve into the role of moderators and their influence on this nuanced dynamic. Leveraging difference-in-differences as the econometric identification strategy, we analyze panel data constructed with chat histories of 7,074 playbacks on Twitch. The results suggest that (a) existing commenters tend to engage less after the increases in group size; (b) the negative engagement effect is the product of mediation effects by way of the increased topic incoherence and emotional polarity of comment (herein referred to as "comment polarity") that decreases engagement; and (c) live streaming channels adopting bot or human moderators can better curb the negative effect, such that bot moderators are effective in decreasing the escalation of incoherence, particularly when the incoming Raider group is large, although human moderators better limit surges in comment polarity, particularly when the incoming Raider group is small. The findings in this paper indicate a congestion effect, a negative externality of increasing group size on commenter engagement in synchronous content platforms, further revealing a nuanced relationship among group size, topic incoherence, comment polarity, and user engagement. Our research further suggests the beneficial role of content moderators, which provides implications for online platform operators and policymakers.

1. Introduction

Synchronous content platforms have become an increasingly prevalent channel among online users for entertainment and information. Live-streaming platforms like Twitch.tv and YouTube Live rank among the fastest-growing and most dominant synchronous content services. Live streaming platforms depend solely on user-generated content (UGC) to captivate audiences. However, distinct from conventional UGC platforms like Twitter, Facebook, and Yelp, where users primarily interact with content through asynchronous actions such as likes and comments, live streaming platforms offer a unique avenue for content creators and viewers to engage in real-time interactions (Figure 1). With 30 million daily active viewers (Shewale 2024) and 51 million hours watched daily (TwitchTracker 2024), Twitch.tv reigns as one of the most popular live streaming platforms. This parallels a major shift as Internet

Figure 1. (Color online) Sample Twitch Live Streaming Channel
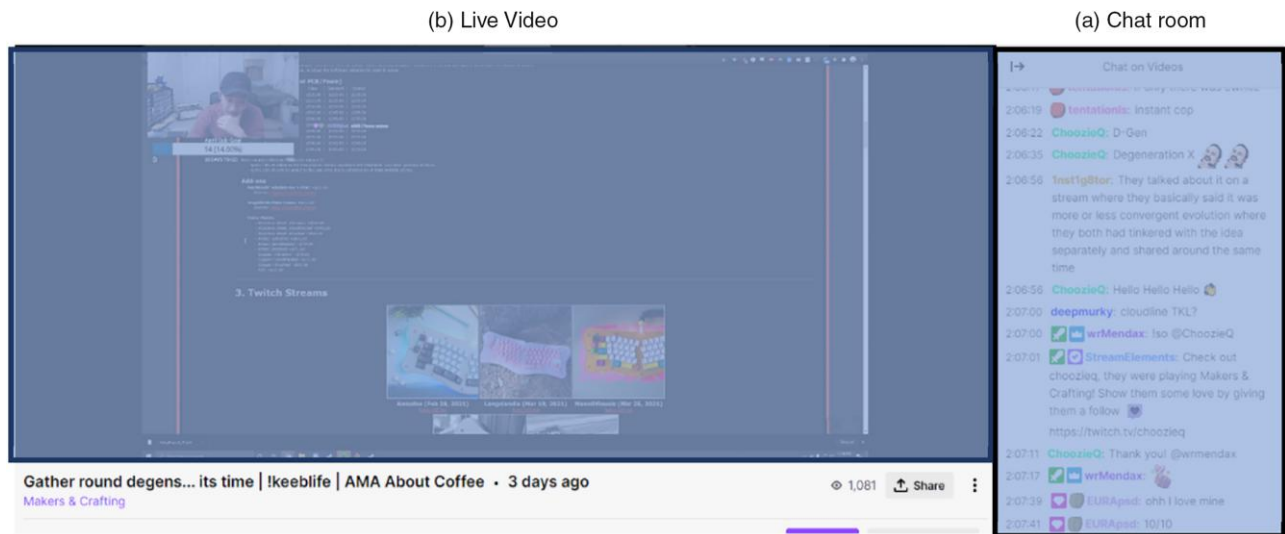
users devote more time to watching live streams than accessing asynchronous content platforms (Dean 2024).

Attracting sizable audiences is imperative for live streaming platforms and streamers alike, as the advertising and subscription revenue largely hinge on the viewer numbers. Increased viewership can also generate a positive network effect, a phenomenon widely seen on social media platforms (Katz and Shapiro 1985, De Reuver et al. 2018, Zhao et al. 2023). A user derives greater utility from a content piece when a larger audience engages with it, which, in turn, boosts the likelihood of further engagement with the content. The chat room design in live streaming proves pivotal in catalyzing this network effect. On platforms like Twitch, viewers can communicate with each other and the streamer via text comments in a chat room (Figure 1(a)) while watching the streamed video (Figure 1(b)). Prior research has shown commenting in the chat room to be a critical part of the live streaming experience, as it enables viewers to feel connected to the streamer and other viewers (Hamilton et al. 2014, Chen and Lin 2018). Stimulating and managing coherent and balanced conversations in the chat room is highly important, as it facilitates enjoyment in watching live streaming and fosters social relationships between viewers, which indirectly drive revenue through gifting and purchases (Wulf et al. 2018, Guan et al. 2022, Zhang et al. 2023).

However, the surge in synchronous content creation and consumption on live streaming platforms, especially with large audiences, comes with a downside. This proliferation leads to a negative network externality commonly known as platform congestion (Huang et al. 2022). The deluge of real-time comments results in information congestion and cognitive overload for users (Schick et al. 1990, Blair 2011). On traditional social

media platforms, the asynchronous approach to content generation and consumption means users can engage with past content at their own pace. When reading an extensive post with numerous comments on Reddit, a user may take several minutes or even hours to fully comprehend the content. Even when a compelling tweet garners thousands of replies, there remains the option to browse through them at a leisurely pace. Additionally, many social media platforms feature recommendation systems, guiding users to the most popular or pertinent content. In contrast, live streaming platforms operate in real time. When a chat channel is deluged with comments in quick succession, viewers find it increasingly difficult to process the sheer volume of information, complicating their efforts to engage in substantive discussions. Much like a conference call overwhelmed by hundreds of participants talking at once on Zoom, excessive simultaneous interactions on streaming channels can disrupt the flow. This makes it harder for streamers and viewers alike to maintain focus and sustain their conversation (Chen et al. 2009). Subsequently, the quality of the viewer experience may be compromised, consequently impacting the financial outcomes for both streamers and the live streaming platforms they operate on. Given these two contradicting forces, our study leverages exogenous group size increases, which provide a unique context for examining these dynamics in real-time environments.

Platform congestion in our context also manifests in the incoherence of topics in an ongoing, multiparty chat. The increase in group size may easily divaricate the current topic of the real-time comments, which makes it even more challenging for streamers and viewers to follow and react. For example, it would be more difficult to respond to 100 comments on 100 different topics than to

respond to the same number of comments covering only one topic, particularly in live discussions. As a result, the increase in topic incoherence potentially suppresses engagement from active viewers (namely "commenters" thereafter) and introduces chaos to chat rooms, further necessitating interventions via means like content moderation. The other challenge introduced by larger group sizes is increased emotional intensity in comments. The rapid influx of diverse users may create greater risks of emotionally polarized expressions (herein referred to as "comment polarity"), which may discourage participation. Prior literature highlights the impact of emotionally intense comments on online platforms and their effects on discussion quality and user experience (Chmiel et al. 2011, Kramer et al. 2014), which further motivates us to empirically examine the role of comment polarity associated with the group size.

To summarize, in the context of synchronous platforms such as live streaming, although a larger viewer base can help better engage existing commenters due to the positive network effect, their presence on the platform may also be suppressed because of information overload, topic incoherence, and comment polarity. Bearing the above discussion in mind, we propose our first research question: *How does the group size affect commenters' quantity and quality of engagement in synchronous content platforms?*

Although understanding how group size impacts engagement provides initial insights, mitigating negative effects poses critical governance challenges. As audiences scale rapidly on live streaming platforms, community norms become harder to maintain, and risks of emotionally charged behavior increase. This drives the need for moderation policies and tools to balance participation incentives against safety. For example, in February 2021, one of the popular Twitch streamers was banned due to a lack of moderation over commenters' misbehavior,[1] attesting to the need of comment moderation for streamers. This is consistent with our prior discussion, where live streams can produce a massive number of comments, leading to high topic incoherence and emotional intensity that obstruct communication. Thus, from a practical standpoint, it's imperative for platforms to manage these comments adeptly.

Studying comment moderation carries theoretical implications regarding synergies between participation and governance. The size of a group amplifies the tensions between network effects and congestion, which can have nuanced implications for the sustainable growth of a community. Examining how streamers navigate these dynamics sheds light on the interplay between scaling incentives and governance capabilities. This informs scalable platform designs capable of sustaining user engagement. Practically, most live streaming platforms have introduced certain types of moderators to regulate real-time communication in the

chat room. Moderators perform the duty to audit and exclude comments against community rules during live streaming, such as spam, hate speech, and scams. Moderators also engage general viewers with community-building activities such as creating polls, broadcast notifications, and background introductions. Streamers can choose to recruit viewers as (human) moderators or adopt algorithm-based (bot) moderators to help regulate viewers' comments.

Particularly, bot moderators have become increasingly popular on live streaming platforms. To integrate bot moderators into the chat room, streamers grant authorization to the third-party bot developer and type preinstalled commands to activate functionalities. Although human moderators serve the dual roles of rule enforcer and conversation facilitator (He et al. 2024), bot moderators' actions are dictated by specific criteria set by algorithmic models, such as removing chats by keyword filtering. Although human and bot moderators are increasingly prevalent on live streaming platforms, their efficacies in managing topic incoherence and comment polarity in synchronous online communication as group size increases remain unclear. Thus, we seek to address our second research question: *How does the presence of human and bot moderators play a role in the group size effects in synchronous content platforms?*

To address these research questions, we leverage the exogenous group size shocks from the *Raid* functionality and the variations in the presence of human and bot moderators on Twitch. On Twitch, Raid is a function that allows a streamer to direct viewers to another ongoing streaming channel at the end of a streaming session. For example, when Streamer A is about to end a broadcast, the option to send a Raid request directs viewers to seamlessly switch to Streamer B's channel. Once a viewer in A's channel accepts the Raid request, the viewer will be automatically transferred into streamer B's channel. From the perspective of streamer B, particularly B's existing viewers, Raid is unexpected, and it exogenously increases streamer B's viewer size. We herein refer to Streamer B's channel as the *Raided channel*. This unique feature offers us a valuable opportunity to identify the effect of group size in synchronous content platforms. Further, given the variations in the presence of human and bot moderators across different channels, we are also able to estimate the efficacy of different moderators in managing real-time engagements when group size increases.

We collected chat comments, complete with time-stamps and channel information, from live streaming playbacks on Twitch between May 15 and September 23, 2022. This data set includes all video playbacks across the top 10 categories on Twitch. We identify the Raided playbacks and match them with identical playbacks without Raid from the same channel. We

aggregate chat comments at one-minute intervals to create a panel data set based on the resulting 7,074 playbacks. Using state-of-the-art deep learning methods, we further construct topic incoherence and comment polarity measures, respectively, for each one-minute interval in each playback. We then employ difference-in-differences models to empirically identify the effect of Raid and the moderating effects of human/bot moderators. We next examined the mediation effects of topic incoherence and comment polarity through mediation analyses. The key results from our empirical estimations suggest that (a) viewers tend to post fewer comments, and there are fewer viewers willing to post comments after the increase in group size; (b) the negative effect of increased group size on existing viewers' engagement is partially mediated by increases in topic incoherence and comment polarity; and (c) whereas both bot and human moderators have the potential to curb the negative group size effect, bot moderators are most effective in decreasing the escalation of incoherence, particularly when the incoming Raider group is large, and human moderators are better at limiting surges in comment polarity, particularly when the incoming Raider group size is small.

This study contributes to prior literature in three ways. First, we contribute to the user engagement literature on digital platforms. Although prior seminal papers have delved into the influence of group size in asynchronous content platforms, such as Wikipedia contributions (Zhang and Zhu 2011), blog posts (Yang et al. 2019), and Tweets (Toubia and Stephen 2013), our research explores the group size effect on user engagement within synchronous environments. Although extant studies show that larger user groups on digital platforms typically boost user contributions due to the allure of social reputation gains, our findings underscore an underexplored facet, suggesting that an expansion in group size can hamper pre-existing user engagement in synchronous, real-time settings, partially through the increased topic incoherence and comment polarity. Second, our study contributes to the online content moderation literature by comparing the efficacy of moderators on UGC platforms (Ruckenstein and Turunen 2020, He et al. 2024). Through the empirical evidence reported in this research, we show that adopting either bot or human moderators can help live streaming channels curb the negative group size effect while each excels in different areas. Finally, our research showcases platform congestion as a negative network externality in the context of synchronous online content platforms (e.g., live streaming). As group sizes increase in real-time settings, they bring forth challenges in decreased coherence and increased polarity. We also emphasize the role of content moderators in maintaining order and decorum, particularly when large groups of viewers convene on live platforms.

2. Theoretical Background
2.1. Synchronous Communication, Chat Comments, and User Engagement

With the significant expansion of Internet bandwidth, various digital platforms have emerged to facilitate content sharing among users, synchronously or asynchronously. Unlike asynchronous platforms such as Twitter and Wikipedia, in which users are allowed to consume content at any time, synchronous digital platforms are constrained by content timeliness. On most platforms that provide synchronous services, such as live chat and live streaming, users can only consume content and interact with peers in real time. For example, customers can ask live assistance questions in a live chat, which will result in a significantly shorter waiting time, fewer instances of information discontinuity, and more precise responses. This unique "live" nature of communication has proven to be effective in terms of reducing the perceived risks in e-commerce (Hong et al. 2021), building interpersonal trust (Qiu and Benbasat 2005), and increasing product sales (Tan et al. 2019). This distinct feature is also applied to fulfill users' social needs, such as tension release (Sjöblom and Hamari 2016) and knowledge-seeking (Kaytoue et al. 2012).

Live streaming, as the most representative service among synchronous platforms, has drawn tremendous attention globally. Live streaming platform Twitch has witnessed a growth of 101%, with 1.7 billion hours watched per month by May 2020 (Stephen 2020), making Twitch surpass Twitter in terms of the total time spent on the platform. Clubhouse, another live streaming platform launched as a live conference social application in 2020, has quickly acquired 10 million weekly active users by December 2020 (Dean 2021). The success of live streaming platforms and streamers critically depends on viewer engagement. Compared with traditional social media, live streaming platforms exhibit fundamentally different user engagement patterns because of their synchronous interactions. Specifically, the real-time nature of live streaming emphasizes the importance of timeliness in user engagement. For example, comments posted 10 minutes earlier might be overlooked by current viewers as viewers may experience an information discontinuity (i.e., a topic no longer relevant) or those comments have been crowded out by others (Christanto et al. 2011). Also, because of the emphasis on real-time interaction, viewers' comments tend to be shorter and more casual than those on asynchronous content platforms, such as Yelp or Wikipedia. This is further exacerbated for streaming channels with a large number of viewers. This is because viewers in these popular channels can be easily distracted or even overwhelmed by a large number of comments posted just within seconds. Considering a user on the platform as a recipient of information, we believe a certain

bandwidth of the information perception limits the comments they may perceive (Blair 2011). Given the interactive needs of live streaming viewers (Sjöblom and Hamari 2016), a negative network externality of information potentially prevents viewers from continuous engagement.

Among various types of engagement on live streaming platforms, chat comments stand out as critically important channels, distinguishing the experience on live streaming platforms from that of prerecorded video and enabling real-time social interactions among viewers and the streamer (Hamilton et al. 2014, Zhao et al. 2021). Prior work highlights how commenting fosters community engagement and relationships. Chen and Lin (2018) find that chatting enables viewers to feel connected to streamers and other viewers and argue that seamless conversations enhance user enjoyment and stickiness. Furthermore, chat activities drive multiple revenue streams on live streaming platforms, including gifting and subscriptions (Wulf et al. 2018, Zhang et al. 2023). Even in the absence of direct purchases, commenting can bolster social perceptions, thereby incentivizing gifting (Li and Guo 2021). More broadly, communication research suggests that coherent, nonextreme discussions are vital to positive user experiences in online communities (Cheng et al. 2017). Hence, the volume and quality of chat comments represent an important form of participation and value creation on live streaming platforms.

2.2. Group Size Effect on User Engagement

Extensive research highlights the impact of group size on motivating individual contributions across various UGC platforms. A seminal paper by Zhang and Zhu (2011) explores the relationship between group size and incentives to contribute by examining the change in the contribution level of users in nonblocked areas when Chinese Wikipedia in mainland China was blocked. Their study revealed that as contributors' social benefits decreased due to the reduced group size, they reduced their content contribution accordingly. Similarly, a recent field experiment on online forum size indicates the same positive effect of group size on individual engagement (Baek and Shore 2020). Furthermore, Toubia and Stephen (2013) found that image-related utility became larger for most users when synthetic followers were added to these users. Guo et al. (2020) enriched the literature by showing the positive effect of virtual crowdedness on user contribution in Waze, a geo-segmented and crowdsourced social mobile app.

Prior studies on group size effects primarily focus on asynchronous platforms, whereas live streaming differs from conventional UGC due to its synchronous interaction between streamers and viewers, and among viewers. As such, it is theoretically unclear whether the group size effect observed in traditional asynchronous content platforms can be generalized to live streaming platforms. One related paper by Lu et al. (2021) examined the live streaming context and showed that adding synthetic viewers in live streaming increases viewer count and average tip per viewer. However, the synthetic viewers added to streaming channels did not increase interaction (i.e., post any comments in chat). This is significantly different from real-world scenarios in which additional viewers may interact through commenting. Exploring the Raid feature in Twitch, which represents an exogenous increase in active viewers, our research also diverges from Lu et al. (2021) to focus the effect of group size on user engagement rather than on tips. Below, we theorize two competing forces arising from the increase in group size in live streaming channels.

On one hand, the theory of network effects indicates a positive effect of increasing group size on overall engagement. In a communications network, the presence of users' adoption behaviors affects other users' participant decisions (Katz and Shapiro 1994) because individuals can derive higher utility if they interact with a larger group of peers (Zhang and Zhu 2011, Toubia and Stephen 2013). A natural extension of this line of work to live streaming is that an increasing active viewer size may enhance interactivity within the chat, leading to a higher level of presence and a higher likelihood that a chat comment is reciprocated. On the other hand, an increase in the active viewer size may lead to a negative network externality: congestion. As many viewers start to engage in the chat, the processing time for each chat comment diminishes quickly. This is further compounded by the real-time nature of the setting, such that the viewers are required to process the information in the moment. When a larger group of viewers interact in a live streaming session, there is a higher chance of more comments being posted in a short period of time, which, in turn, increases information congestion and leads to a drop in response rates due to the input surpassing the limits of cognitive capacity (Schick et al. 1990). As such, existing viewers' willingness to participate in chats decreases, and engagement level lowers.

More specifically, the negative network externality arising from congestion encompasses both quality and quantity dimensions (Simpson and Prusak 1995). From a quantity perspective, congestion means overload, which refers to receiving too much information within a certain time period (Schick et al. 1990, Chen et al. 2009). For example, in a live streaming context, the quantity of information can be operationalized as the number of comments posted within a period of time (e.g., a minute). Meanwhile, congestion also has a quality component: an influx of comments that lack substantive contribution to the discussion exacerbates the congestion. In the event of congestion, content submissions may become erratic, losing coherence, potentially containing

very emotional expressions, and offering limited value or disruptive experiences to viewers, irrespective of the quantity (Speier et al. 1999, Jones et al. 2004). Thus, it is important to also evaluate the quality implications of congestion. In this study, we consider two quality measures based on previous literature: topic incoherence (Speier et al. 1999) and comment polarity (Lyu and Kim 2016). These two quality measures align with the cognitive-emotional processing theory (Pessoa 2008), where topic incoherence reflects cognitive processing challenges, and comment (sentiment) polarity corresponds to the emotional intensity of the information environment.

Cognitively, an increase in group size potentially raises the risk of topic incoherence, which can discourage participation. As the number of participants grows, the diversity of topics and perspectives increases (Salehabadi et al. 2022), creating an environment that challenges viewers' ability to follow and contribute to coherent discussions. In other words, the influx of new viewers may introduce a greater variety of subjects, making it more difficult to maintain a focused conversation. Consequently, existing users are more likely to perceive discussion to be incoherent (Kumar et al. 2010). Because viewers' attention is a scarce cognitive resource, when new viewers may not have common knowledge as existing viewers, existing viewers have a great chance of being diverged by the different topics of comments posted by other incoming users (Adamczyk and Bailey 2004).

Emotionally, larger group sizes can lead to more polarized comments, further impacting engagement. Prior research has shown that extremely positive or negative comments can prompt individuals to disengage from online discussions (Kramer et al. 2014, Cheng et al. 2017). As the number of participants grows, the likelihood of encountering emotionally charged comments increases, potentially creating an environment that undermines meaningful conversations. Exposure to such polarized content is shown to reduce user enjoyment and lead to less engagement (Mohan et al. 2017). In other words, the influx of new viewers may introduce more polarized emotional expressions, making it more challenging to maintain a balanced emotional climate in the chat. In sum, with a larger audience size, users are more likely to experience disutility from diverse topics and emotional strain from polarized comments, creating additional challenges for existing viewers to continue their engagement. Beyond these dimensions, it is also important to consider the nature of the group size effect, as changes in group size can influence engagement through other mechanisms, including interruptions. Online users are typically aware of changes in group size (e.g., through notifications or visible metrics), and such changes inherently create a sense of discontinuity for existing participants. This discontinuity, which is

part of the quantitative aspect of group size, may exacerbate congestion effects by disrupting the flow of discussion and altering the interaction dynamics (Adamczyk and Bailey 2004).

In the context of live streaming, where real-time interactions amplify the effects of content characteristics, understanding how group size increases affect both the quantity and quality of engagement becomes crucial. As platforms and content creators aim to maximize engagement, they must also consider the potential negative effects of information overload, topic incoherence, and comment polarity that may arise from larger group sizes. To sum up, based on the discussion above, we seek to empirically evaluate (a) *the group size effect, whether the increase in group size has a negative or positive effect on the engagement of existing active viewers*; and (b) *potential mechanisms, whether increased topic incoherence and comment polarity serve as mediators.*

2.3. Bot and Human Moderators on Digital Platforms

Streamers on Twitch widely adopt two types of moderators: bot moderators and human moderators. Bot moderators are algorithm-based programs that perform automated tasks, whereas human moderators are mainly users who volunteered to help manage the channel. Moderators have two primary methods to encourage and regulate viewer engagement, that is, posting messages in chat or banning users who violate channel rules (Seering et al. 2018, Seering and Kairam 2024).

The influence of bot moderators on online community regulation has started to receive attention in recent literature. For example, Seering et al. (2018) show the effectiveness of proactive moderation tools in diminishing certain content posting, whereas reactive bans discourage a wider variety of behaviors. Matias (2019) conducts a randomized experiment on a science discussion community and finds that announcing community rules frequently motivates newcomers to participate in the discussion. Srinivasan et al. (2019) examine the change in users' future engagement, achievement, and compliance after comment deletion and reveal that removing problematic comments leads to an immediate decrease in noncompliance rates. Bot moderators efficiently handle repetitive tasks that involve handling negativity and making announcements, which creates an informative and healthy atmosphere for the online community, thus promoting communication among participants.

Human moderators play an important role in online communities. As a general rule, human moderators are mostly experienced community members volunteering to help regulate the community. Their main responsibilities include but are not limited to answering questions, deleting rude messages, and banning commenters (Wohn 2019). Seering et al. (2019) interviewed 56

volunteer moderators (i.e., human moderators) to determine how moderators engage with their communities. They emphasize the importance of protecting user motivation and finding a balance between intervention and protection of viewers' right to speak. Compared with bot moderators, human moderators serve as managers for the community and perform more corrective and supportive activities (He et al. 2024), which may lead to more productive and efficient communication. Extant studies explore the coworking mechanism of the bot and human moderators and examine their role in regulating activities and motivating engagement. He et al. (2024) find that human moderators' role as community managers is augmented when moderation tasks are delegated to machines (bot moderators).

This study aims to examine the moderating roles of bot and human moderators on the impact of group size in a synchronous communication platform, a gap in previous literature. In the context of changing group size, human moderators can quickly adapt to shifts in conversation and employ their judgment to guide participants. They comprehend subtle changes in tone, humor, and intent that might lead a discussion astray, allowing them to intervene appropriately and maintain viewer engagement (Seering et al. 2019). This adaptability is crucial in managing the dynamic nature of live streaming chats, where conversation topics and participant behavior can change rapidly (Wohn 2019). Additionally, human moderators can resort to functions such as "slow mode" when chat becomes overwhelming, effectively slowing down the chat and restoring balance. These unique characteristics could be particularly crucial when new users enter the community and need to grasp the norms and background, especially when the incoming group is small. This aligns with findings from Cai and Wohn (2022), who highlight the importance of human moderators in maintaining community norms and guiding newcomers.

On the other hand, bot moderators manage the chatroom in a different manner. Instead of trying to comprehend the chats and perform manual interventions, bot moderators (e.g., NightBot) can perform scalable actions such as sending announcements and automatically removing spam. Bot moderators operate based on predefined rules and algorithms, which allows them to consistently enforce channel norms and maintain structure in the chat (Seering et al. 2018). By quickly removing off-topic or spam messages, bots can create a more focused environment, potentially helping viewers to follow and engage with the main discussion topics more easily (Matias 2019). Unlike human moderators, who may be influenced by subjective factors, bot moderators can also provide more consistent moderation across varying group sizes. Additionally, regular announcements from bots can serve as reminders of current topics and channel norms, which may help maintain topic coherence

even as group size increases (Srinivasan et al. 2019). Bot moderators' potential efficiency in handling larger groups aligns with their algorithmic nature, allowing them to process and moderate high volumes of messages quickly when dealing with large influxes of viewers (Seering et al. 2018, Matias 2019).

Notably, although bots may excel at enforcing explicit rules and managing large volumes of messages, they might struggle with the contextual understanding necessary to foster deeper engagement (He et al. 2024). This is consistent with literature that shows bot moderators face challenges in understanding complex social contexts and nuanced communication (Chandrasekharan et al. 2018). Based on the discussion above, although we do not hypothesize the moderated mediation relationships, we seek to explore whether the presence of human moderators effectively attenuates the increase in comment polarity (emotional) arising from larger group sizes, particularly for smaller incoming Raider groups. And we also evaluate whether the presence of bot moderators effectively attenuates the increase in topic incoherence (cognitive) arising from larger group sizes, particularly for larger incoming Raider groups.

3. Data Set and Measurement Construction

3.1. Data Set

We collect our research data from Twitch, the leading live streaming platform worldwide, from May 15 to September 23, 2022. The data collection involves the following steps. We first identify the top 10 categories on Twitch, defined by the platform in terms of viewership, including Just Chatting, Minecraft, Valorant, Counter-Strike: Global Offensive, Call of Duty: Warzone, Fortnite, and Warcraft. Within these categories, only "Just Chatting" is a nongaming related category where streamers usually engage in real-time conversations, discuss various topics, or simply hang out with their viewers. By contrast, other categories focus on specific video games, gameplay strategies, and interactive sessions between the streamer and their audience. We capture the streamer IDs for active streamers in these categories by recording those who are streaming every 30 minutes during the collection period, resulting in a large pool of active streamers who have streamed at least once during our collection period. Through Twitch API, we also captured information of those active streamers by their ID. Streamers' information contains the streamer ID, name, number of followers, creating time, and the partnership with Twitch.

Upon generating the streamer pool, we recorded all video IDs associated with the collected streamers. Notably, Twitch automatically archives streamers' most recent live streaming (i.e., for the past 60 days) by uploading the playback along with corresponding chat

history. Thus, we are able to access the comments posted by the viewers and locate each comment with a timestamp. We then captured a detailed history of comments following each video ID. Examples are provided in Figure 1(a). For each comment, we can observe the commenter's name, role, comment text, and comment timestamp (i.e., seconds to the start point) accordingly. It is worth mentioning that the system notification is also included in the comments area, enabling us to identify the timestamp of the Raid and the number of Raiders (i.e., viewers who change to the Raided channel because of the Raid function). We then limit the language of the playback to English and identify the Raid notification based on the role of commenters and notification text. Raid announcement is shared through a system notification following a pattern of "X Raiders from Y have joined!," where X represents the number of viewers who participate in the Raid and Y denotes the source streamer who initiates the Raid. Next, we mark playbacks with Raids according to our objectives. The resulting data set contains 129,119 Raid records from 6,611 streamers.

Finally, we created a panel data set based on the streamer, playback, and chat records collected in previous steps. We locate the timestamp of each Raid and calculate the time distance in minutes for each comment. Then, we use negative integers for comments made before the Raid and positive integers (starting from zero) for comments made after the Raid to incrementally index the time distance as our observational sessions. For example, one comment posted at 30'45'' in the playback is indexed as session -1 if the Raid happens at 31'10''. Similarly, a comment posted at 31'12'' would be indexed as session 0. Next, we aggregate the chat comment records by sessions and calculate the engagement measure. For example, to calculate comments per commenter for a session by streamer i in period t, we divided the number of comments in this session by the number of commenters within the same session. We further mark the moderator type at the video level, as streamers do not change their moderator type within one playback. Dummy variable *Human* equals one if human moderators are identified in the video, and *Bot* equals one if bot moderators are identified; otherwise, they equal zero. We identify bot moderators by matching any moderator in the chat history to a comprehensive Twitch bot list. As discussed above, our estimation target is the group of viewers who have engaged before the Raid (existing commenters). Following our definition that existing commenters are the audience who has engaged in the channel before the Raid, we operationalize this definition if one viewer has posted a comment before the Raid point.

To further mitigate potential endogeneity concerns due to Raid selection biases, we utilize both exact matching and propensity score matching (PSM) to create a comparable control group of playbacks without Raids. It is notable that after we identify the "treatment" group as playbacks that experienced a Raid, we first conduct exact matching to narrow down the control group playback pool on two constraints: the playback from the *same streamer* in the treatment group, with the same day of the week (e.g., Monday), and same time of the day (e.g., Morning). In other words, our exact matching imposes that (a) the playbacks being matched on are from the same channel, and (b) the playback's publishing times for the matched sample are within the same time of the day and day of the week. This approach allows us to generate high-quality control group playbacks as counterfactuals.

Based on the resulting data set from exact matching, we next perform the PSM to match the playback and streamer characteristics that may influence engagement patterns by first estimating propensity scores − the probability of receiving the treatment conditional on observed covariates: the cumulative views for the video and log-transformed Unix timestamp of the video publishing date as the covariates. These covariates allow us to control for the popularity and recency of the videos, which could influence engagement patterns. We use a nearest neighbor propensity score approach with calipers of 0.2 standard deviations of the logit of the propensity score. Based on standardized mean differences (SMDs) before and after matching, there is substantial improvement in comparability and matched sample SMDs across covariates are within the 0.1 thresholds (close to zero). Together with the matching criteria above, playbacks in the treatment group are highly comparable with those in the control group. The matching performance is reported in Online Appendix 1, Section A1.1.

In addition, we mark the relative time for the matched control group by the interval from the starting time of the playback. Consider playback A in the control group matches with playback B in the treatment group, if a Raid happens at 30'00'' in playback B, we mark the 30'00'' in playback A as the cutoff point as the counterfactual Raid time. This approach results in a matched sample where the treatment and control groups have comparable prior distributions on unobserved dynamic factors related to stream flow. Overall, the combination of exact matching and PSM ensures the comparability of the treatment and control playbacks, which allows us to attempt a causal interpretation of the findings.

3.2. Measuring Engagement

Depending on the specific platform being studied, previous literature employs various measures for user engagement on asynchronous platforms. In general, the measures are typically at the levels of the users (Kuang et al. 2019), comments (Zhang and Zhu 2011), or content (Yang et al. 2019). Similarly, the emerging literature on

synchronous platforms primarily measures engagement through the number of users or number of contents (i.e., in our context, comments). We present a summary of engagement measures in Online Appendix 1. In this paper, we focus on commenting engagement, and construct the following four measures to capture the engagement for both quantitative (extensive and intensive margins of commenting) and qualitative (cognitive and emotional processes in comments) perspectives.

Number of Commenters measures how many viewers engage in the live commenting by counting distinct usernames within each session. This measure intuitively denotes the size of active viewers, which is used by existing live streaming literature to capture active participation (Jones et al. 2004) at the *extensive* margin. *Comments per Commenter* measures count of comments posted by existing commenters divided by the number of existing commenters within one session. It captures the *intensive* margin of user engagement. This measure has been used to evaluate the user contribution in the context of Wikipedia (Zhang and Zhu 2011). *Topic Incoherence* measures average semantic distance between comment topics. This measure is calculated based on all comments within each session (one minute). *Comment Polarity* measures the absolute value of sentiment scores detected by a state-of-the-art sentiment analysis model. This measure is calculated based on all comments within each session. We next discuss the construction of topic incoherence and comment polarity in more details.

3.3. Topic Incoherence Measure

Following the suggestion from Yang et al. (2019), we seek to explore the effect of group size on the content features of comments. Specifically, we are interested in the change in topic consistency level caused by the Raid. In linguistics, the term "discourse incoherence" refers to a lack of logical integration in information, resulting in disconnected or inconsistent communication (Foltz et al. 1998). Incoherence has been widely used as an important measurement of textual information quality. A wealth of research has evidenced that a less incoherent discussion environment can better reduce participants' inference load during comprehension (Grosz et al. 1995). On live streaming platforms where chats are massive and convulsive, the incoherence of chat messages is considered a key factor contributing to user engagement. Existing research shows that communication practices that aim to restrain discourse incoherence, such as contraction of text into a smaller space and adoption of shared viewpoints, can make massive chats legible, meaningful, and compelling to participants (Ford et al. 2017). In the focal context, we refer chat "topic incoherence" to the content-unrelatedness of chat messages posted in the chatroom.

Quantitating the incoherence of a set of statements has received substantial attention in the past few years.

Researchers have developed different methods to measure discourse incoherence, such as the well-established term frequency-inverse document frequency (TF-IDF) (Salton and Buckley 1988) and latent Dirichlet allocation (LDA) (Blei et al. 2003). However, these approaches do not fit the Twitch chat messages for several reasons. First, these methods encode word co-occurrence information only at the document level, which limits their capabilities for short sentences (e.g., Twitch chat messages). Second, unlike structured documents that have countable keywords, the topics of Twitch chat messages are infinite as the content of the videos varies greatly from each other, and so does the way people chat. Moreover, the topics of posted comments in Twitch chat rooms are always changing, and it is impractical to capture the dynamic time trend of chat messages using traditional methods.

To model the topic incoherence through texts, we adopt a neural network-based method called word embedding (Levy and Goldberg 2014) to vectorize comments at the word level. Word embedding is an unsupervised method that aims to find vector representations for each word by examining semantic similarities between different words. It generates real-valued vectors with a countable dimension by learning from the usage of words, enabling the words used in similar ways to have similar embedding vectors in the learned dimension. That is, the more similar the semantic meaning of two words, the closer they are in the learned embedding space. In practice, measuring the similarity between sentences based on their word similarity is a widely adopted method. A wealth of research uses the average vector of words' vectors to represent text and demonstrates its excellent performance in various tasks such as text classification and summarization (Mikolov et al. 2013, Kusner et al. 2015). Another benefit of utilizing the word embedding method is that embedding vectors are derived from large unannotated corpora, and pretrained embeddings can then be implemented in downstream tasks.

Therefore, we incorporate a well-established pretrained word embedding model, GloVe (Pennington et al. 2014), to vectorize comments at the word level. We apply GloVe in this study not only for its excellent performance on text mining related tasks such as word analogy, word similarity, and named entity recognition but also because it provides word vectors trained from social media platforms to represent elaborate textual information. We leverage the GloVe Twitter vector, which is trained on 2 billion Tweets. The suitability of the GloVe model is further attested by the fact that live streaming comments share many similarities with Tweets, as they both frequently use acronyms and cyberspeak words in a casual form (Hu et al. 2013).

The operationalization of topic incoherence modeling is described as follows. First, we perform word-level

tokenization for each comment to extract individual words in the comment. Next, we employ the pretrained word GloVe to vectorize comments at the word level. Then, the average of each word vector is used to calculate the comment vector. Specifically, the comment vector of comment c is mathematically represented as $h_c = \frac{1}{|W|} \sum_{i=1}^{|W|} v_{w_i}$ ($w_i \in c$), where v_{w_i} denotes the vectors of word w_i that belongs to comment c, $|W|$ is the total number of words that appear in comment c.

Next, guided by the centering theory (Grosz et al. 1995) and discourse incoherence in linguistics (Witte and Faigley 1981), we model topic incoherence using the averaged Euclidean distance among comments. Specifically, we calculate the topic incoherence score of each session by averaging the pairwise distances among the comment vectors. Formally, given a series of comment vectors in a session $\{h_{c_1}, h_{c_2}, \ldots, h_{c_N}\}$ obtained from the previous step, the incoherence score for a single session is $Incoherence = \frac{\sum_{i=0}^{N} \sum_{j=0, j \neq i}^{N} dist(h_{c_i}, h_{c_j})}{N(N-1)}$, and $dist(h_{c_i}, h_{c_j}) = \sqrt{\sum_{m=1}^{k} (h_{c_{im}} - h_{c_{jm}})^2}$. Here N is the total number of comments in the session, and h_{c_i}, h_{c_j} represent the comment vectors of comment c_i and c_j, respectively, that appear in the session. Subscript m denotes the index of dimension in the embedding vector. In our context, the pretrained GloVe model returns 50 dimensions, which is the hyperparameter chosen to project the words vector into the semantic space (Pennington et al. 2014). Here, we use Euclidean distance for the distance function $dist(\cdot)$ based on its excellent capacity to capture verbal analogies. Euclidean distance between word vectors is a good proxy for word dissimilarity, as there is a moderately strong positive correlation between two words' co-occurrence shifted pointwise mutual information (csPMI) and their Euclidean distance in vector spaces (Ethayarajh et al. 2018). It is also demonstrated that Euclidean distance outperforms other distance measurements in many relational similarity evaluation tasks compared with human judgments (Chen et al. 2017), leading to its wide adoption in similarity measures in the domain of natural language processing (Singha and Das 2013). We illustrate the process of constructing the topic incoherence score in Figure 2.

The topic incoherence score enables chat content exploration with semantic distances between comments. The larger the incoherence indicator is, the less likely the comments are integrated and of the same topic. An example of the incoherence score in our data set is provided in Table 1.

3.4. Comment Polarity Measure

We discuss the calculation of comment polarity in this section. Prior research has shown that extreme sentiment, whether positive or negative, can influence participation and shape the dynamics of discussions in online communities (Arguello et al. 2006, Chmiel et al. 2011). In the context of online interactions, comment polarity refers to the degree of emotional deviation from neutral sentiment, encompassing both positive and negative extremes. This concept aligns with the cognitive-emotional processing theory (Pessoa 2008), which posits that emotional and cognitive processes are deeply intertwined in human information processing. In online environments, highly polarized comments may serve as strong affective stimuli, potentially influencing the emotional responses of users.

The absolute value of polarity, regardless of its direction, can indicate the intensity of emotional expression in comments. High polarity comments, whether extremely positive or negative, may capture more attention and elicit stronger reactions from other users, thereby affecting engagement patterns in live streaming contexts (Kramer et al. 2014). Therefore, we use advanced natural language processing techniques to measure comment polarity in our Twitch data set. Specifically, we define comment polarity as a measure representing the intensity of sentiment expressed in a comment, regardless of whether it is positive or negative. Note that our use of polarity in this context differs from the concept of political polarity or ideological polarization. Our polarity measure is specifically

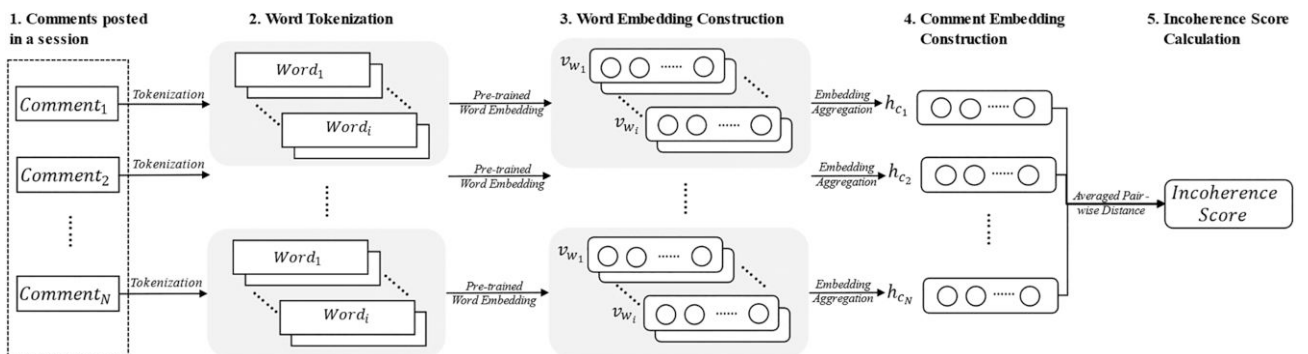**Figure 2.** Procedures to Compute Topic Incoherence Scores

Table 1. An Incoherence Score Example

	Session A	Session B
Message examples	i do that with my fiance all the time	ooo i love maplestory
	Like the meme	I love maplestory too
	Hello im home	i started playing maplestory again
	www	there is a maplestory app now!',
	can you stop watching me	its fun!
	me too	very nice~
	its ok i like hearing Japanese	I know!! I love the app lol
	… ...	… ...
Incoherence score	6.909	2.624

situated within the realm of sentiment analysis, focusing on the intensity of emotional expression in the text. This is widely used in natural language processing and social media analytics (Lyu and Kim 2016, Muhammad et al. 2016). By capturing the degree of emotional intensity in comments, regardless of their valence, we can assess how strongly expressed sentiments might influence user engagement in live streaming environments. Specifically, we leverage a state-of-the-art sentiment analysis model called RoBERTa, developed by researchers at Facebook AI (Liu et al. 2019). This model is fine-tuned on large data sets of social media content to predict the sentiment of the given text. We input Twitch comments within each session through the pretrained RoBERTa model, which outputs a sentiment score, namely *polarity*, ranging from −1 (extremely negative) to 1 (extremely positive). Then, to capture the intensity of sentiment regardless of its direction, we take the absolute value of these polarity scores. This transformation allows us to measure how far the polarity deviates from neutral, with higher values indicating more intense emotional content. By aggregating the average absolute polarity scores at the session level, we obtain time-varying measures of the comment polarity.

This approach enables us to formally measure and test whether increased group size leads to greater polarity (i.e., more intense sentiment expression), which may, in turn, affect engagement among existing viewers. Notably, the RoBERTa model's state-of-the-art performance on sentiment analysis tasks ensures robust assessment of the emotional dynamics in Twitch chat streams. By leveraging these advanced NLP techniques, we can robustly test our hypothesized relationships between group size, polarity, and user engagement. The engagement measures mentioned above for existing viewers are then appended to our panel data. The descriptive statistics of key variables are listed in Table 2.

It is worth noting that the text-based measures (i.e., incoherence and polarity) may contain missing values. The reason for missing values for the incoherence score is that based on our measurement construction, sessions with a single comment or zero comment cannot be used to calculate the incoherence score. As for polarity scores,

sessions with zero comments will also return a missing value. In this work, we fill missing values with zero because a single topic represents high coherence (no incoherence). Also, zero comments have no incoherence or polarity issues. As readers may be concerned that this approach may introduce bias into the estimation, we conduct a robustness check, omitting missing values, and find the results highly consistent.

4. Empirical Analysis
4.1. Main Effect

The Raids in streaming channels provide exogenous variations in group sizes for streamers and existing viewers as they cannot predict the entry of Raiders.[2] This exogenous shock allows us to empirically identify the impact of group size on user engagement. Within stream channels wherein Raid happens, the Raids create a discontinuity in group sizes. Whereas within streams with no Raid, we would observe no discontinuity in group sizes. This empirical setup allows us to employ difference-in-differences (DID) estimations as our identification strategy, which enables us to effectively control the unobservable time trend and reliably estimate the effect of an exogenous increase in group size.

To operationalize our panel structure, we set 20 minutes as the time window, with 10 minutes before and 10 minutes after the Raid. This allows viewers to have sufficient time to react to a Raid, yet this time frame is not so long that the impact of the Raid becomes irrelevant. We believe this selection of time window is sufficient to cover the immediate and relatively long-term effect of one Raid.

Formally, we consider the Raid as the treatment and index data set within a 10-minute time window before and after the Raid as the treatment group. In our empirical setting, the treatment group comprises data sets ranging from session −10 to 9, wherein Raid happens at the beginning of session 0. As discussed in the Data and Measurement Construction, we then define sessions in the matched samples, which are not affected by the Raid, as the control group. The control group comprises a data set ranging from session −10 to 9 with a counterfactual treatment at session 0 (i.e., this counterfactual

Table 2. Descriptive Statistics of Key Variables

Variable	Description	Observations	Mean	Standard deviation	Minimum	Maximum
$Raided_{i,t}$	A dummy variable (0, 1), = 1 if there are Raid viewers for playback i within session t	141,665	0.263	0.444	0	1
$Incoherence_{i,t}$	Incoherence score for all comments in the chat room for playback i within session t. The average Euclidean distance in comments vectors for all viewers (Dissimilarity).	141,665	1.756	1.907	0	10.908
$Polarity_{i,t}$	The absolute value of polarity score for comments for playback i within session t.	141,665	0.27	0.242	0	1
$CommCount_{i,t}$	Commenter count for existing commenter. Number of existing viewers who post comments for playback i within session t	141,665	4.703	14.075	0	545
$CPC_{i,t}$	Comments per existing commenter. The average count of comments for each existing viewer for playback i within session t	141,665	0.66	0.65	0	36
$Human_i$	A dummy variable indicating the presence of human moderator for playback i	141,665	0.899	0.301	0	1
Bot_i	A dummy variable indicating the presence of bot moderator for playback i	141,665	0.852	0.356	0	1
$RaidSize_i$	Number of Raiders joining the playback i	141,665	135.539	2,777.161	0	252,037
$DayOfWeek_t$	A factor indicating the day of week for session t. 0: Sunday; 1: Monday; 2: Tuesday, until 6: Saturday	141,665	2.73	1.994	0	6
$TimeOfDay_t$	A factor indicating the time period in a day. 1: Afternoon; 2: Midnight; 3: Morning, 4: Night	141,665	2.581	1.235	1	4

treatment is identified by the distance to the starting point of the video, as discussed in Section 3). We believe this control group construction is optimal for the following reasons.

First, the control and treatment groups are highly comparable in terms of the streaming time. This setting allows us to control potential viewer engagement heterogeneity across different time of the day and different day of the week. Second, given the unique live streaming culture on Twitch, viewers' engagement patterns are usually unique depending on the streaming content and streamer (Sjöblom and Hamari 2016). Limiting the matched samples to the same channel (streamer) controls for the uniqueness of engagement based on streaming style and audience base. Third, we find that the time trends in the control group and pretreatment samples are highly consistent, satisfying the parallel trend assumption in DID estimation. We present the testing results in Online Appendix 1, Section A1.1.

For sessions in which Raids happen within the first 20 minutes of the playback, we exclude such sessions as they do not have sufficient observations to construct a corresponding control group. Also, there could be multiple Raids in one live streaming video, which may introduce an overlap between the control group for the second Raid and the treatment group for the first Raid. To ensure reliability, we only consider the first Raid in our DID setting. We argue that the aforementioned data exclusions will not affect our results as only 8.2% of the data set was omitted, and the results without excluding these observations are highly consistent with our main results.

As the viewer engagement may be subject to streamers' and playbacks' characteristics and experience, we add a playback level fixed effect in our model. Notably, because streamers can be identified by the playback (i.e., each playback has a unique ID associated with one streamer), the inclusion of the playback level fixed effect will automatically absorb the streamer (channel) level fixed effect, accounting for both time-invariant and some time-varying streamer characteristics, such as the tenure of streamers, channel category, the popularity of streamers and streaming date. In addition, we include fixed effects for day of the week and time of day to control for potential time patterns of how viewer engagement evolves over time. Finally, we add a relative time fixed effect to control for pre-existing trends and to ensure that the estimated treatment effect is not confounded with unobserved time-varying factors. Note that our study focuses on the average effect of raid on viewer engagement rather than to quantify the impact of raider size on viewer engagement, which requires a different econometric specification. We include analyses related to raider size and nonlinear effects in Online Appendix 2.

Relying on the empirical model, we run fixed effect estimation for existing viewers. Given the skewness in the dependent variables, we apply log transformation, which naturally offers percentage interpretations. We estimate the following DID model to examine the

average effect of the Raid:

$$ln(DV_{i,t}) = \beta_0 + \beta_1 Raided_{i,t} + Playback_i + \theta_t + \eta_t + \nu_t + e_{i,t}, \tag{1}$$

where i indexes the playback and t indexes the observational session; $DV_{i,t}$ denotes the dependent variables proposed in the previous sections (i.e., number of existing commenters and comments per commenter). We apply log transformation to the dependent variables (DVs) due to their skewness. $Playback_i, \theta_t, \eta_t, \nu_t$ represent fixed effects at the playback, day of the week, time of the day, and session level. $Raided_{i,t}$ is a dummy variable marked for the existence of Raiders in session t. It is worth noting that, the construct of $Raided_{i,t}$ is equivalent to $After_t \times Treat_i$ where $After$ refers to whether the session belongs to the post-Raid period, and $Treat$ is a dummy indicator referring the Raided playbacks. The main effects of the $After$ and $Treat$ are subsumed by respective fixed effects at the time and playback levels. In this model, the coefficient β_1 captures the average effect of the Raid, $e_{i,t}$ denotes the error term.

4.2. Mediation Analyses

To investigate the underlying mechanisms of the group size effect, we employ a causal mediation analysis using a parallel dual-mediator model. This approach allows us to formally test the hypothesized roles of two key constructs: comment incoherence and comment polarity. Our methodology is grounded in recent advancements in causal inference and mediation analysis (Valeri and VanderWeele 2013, VanderWeele 2016), which extend beyond the traditional approach of Baron and Kenny (1986) by incorporating exposure-mediator interactions and allowing for a clear causal interpretation based on counterfactuals.

We implement a two-stage regression-based procedure for the model under a panel data structure, following the framework outlined by Bai et al. (2020) and Li et al. (2022). In the first stage, we estimate the effect of our treatment on each mediator (incoherence and polarity) using Equation (2-1) and (2-2). In the second stage, we estimate the effects of the treatment and mediators on our outcome variables using Equation (2-3). This approach allows us to calculate the natural indirect effect (NIE) for each mediator and the natural direct effect (NDE) of the treatment.

Specifically, the NIE quantifies how much the outcome would change if the Raid treatment was held constant, but the mediator changed from its natural level under no treatment to its level under treatment. The NDE captures how much the outcome would change if the treatment changed but the mediator remained at its natural level without treatment. By calculating these effects, we can estimate the proportion of the total effect that is mediated through each pathway. Formally,

following a similar setup in Li et al. (2022), Hayes (2017), and Bai et al. (2020), our estimation consists of two steps as follows:

Step 1:

$$Incoherence_{i,t} = \beta_{10} + \beta_{11} Raided_{i,t} + Playback_i + \theta_t + \eta_t + \nu_t + e_{i,t}, \tag{2-1}$$

$$Polarity_{i,t} = \beta_{20} + \beta_{21} Raided_{i,t} + Playback_i + \theta_t + \eta_t + \nu_t + e_{i,t}. \tag{2-2}$$

Step 2:

$$ln(DV_{i,t}) = \beta_{30} + \beta_{31} Raided_{i,t} + \beta_{32} Incoherence_{i,t}$$
$$+ \beta_{33} Polarity_{i,t} + \beta_{34} Raided \times Incoherence_{i,t}$$
$$+ \beta_{35} Raided \times Polarity_{i,t} + Playback_i + \theta_t + \eta_t$$
$$+ \nu_t + e_{i,t}, \tag{2-3}$$

where $Incoherence_{i,t}$ is the incoherence score we construct to measure the semantic dissimilarity within one minute of live comments, which is indexed by session t and playback i. $Polarity_{i,t}$ is the comment polarity probability within session t for playback i, predicted by the pretrained model. Other notations remain the same as ones in Model (1).

Using this two-stage approach, we can quantify the NIE and NDE in the relationship between group size and engagement. For continuous outcomes and mediators, we are interested in interpreting NIE and NDE (VanderWeele 2016, Li et al. 2022), which are calculated as NIE_Incoherence = $(\beta_{32}\beta_{11} + \beta_{34}\beta_{11}a)(a-a^*)$; NIE_Polarity = $(\beta_{33}\beta_{21} + \beta_{35}\beta_{21}a)(a-a^*)$; NDE_Incoherence = $\{\beta_{31} + \beta_{34}(\beta_{10} + \beta_{11}a^*)\}(a-a^*)$; and NDE_Polarity = $\{\beta_{31} + \beta_{35}(\beta_{20} + \beta_{21}a^*)\}(a-a^*)$. Here a represents the treatment condition (if being Raided, $a = 1$, otherwise 0), and a^* represents the control condition (if not being Raided, $a^* = 1$, otherwise 0). Then, to further compare the indirect effect, we are interested in the proportion mediated by each mediator, calculated as NIE/(NIE + NDE). This allows for a clear causal interpretation of the mediation results based on counterfactuals for each mediator separately, as outlined in recent causal mediation literature (Valeri and VanderWeele 2013, VanderWeele 2016, Li et al. 2022). Significant indirect effects would provide evidence that changes in incoherence and polarity are mechanisms driving the change in participation when group size increases unexpectedly. We present more detailed discussions in Online Appendix 4.

4.3. Effect of the Presence of Moderators

We further investigate the moderating roles of human and bot moderators. Specifically, we allow the Raid treatment effects on the mediators (incoherence and polarity) to vary depending on moderator presence. This involves interacting the Raid indicator with binary indicators for human and bot moderators and

estimating the first step separately for each moderation condition. In other words, we add *Human* and *Bot* into the interaction term of the main effect to examine their moderating effects for Equations (2-1) and (2-2). Because the variable *Human* is binary, this estimation uses observations in which *Human* = 0 as the baseline. Note that the baseline is the scenario in which there are no corresponding moderators. We evaluate whether and how human versus bot moderators attenuate the relationship between Raids and engagement by managing comment incoherence and polarity. We predict that the presence of moderators will partially counteract the surges in incoherence and polarity arising from larger group sizes.

Specifically, we estimate the equations as follows:

$$Incoherence_{i,t} = \beta_{10} + \beta_{11} Raided_{i,t} + \beta_{12} Raided_{i,t} \times Human_i$$
$$+ \beta_{13} Raided_{i,t} \times Bot_i + Playback_i + \theta_t + \eta_t$$
$$+ \nu_t + e_{i,t}, \tag{3-1}$$

$$Polarity_{i,t} = \beta_{20} + \beta_{21} Raided_{i,t} + \beta_{22} Raided_{i,t} \times Human_i$$
$$+ \beta_{23} Raided_{i,t} \times Bot_i + Playback_i + \theta_t + \eta_t$$
$$+ \nu_t + e_{i,t}, \tag{3-2}$$

where $Human_i$ is an indicator marked at the playback level for the presence of human moderators. Similarity, the presence of bot moderators is marked as Bot_i. We do not add the main effects of the presence of mediators because it would be absorbed by the playback fixed effect.

5. Results
5.1. Group Size Effects
To examine the direct impact of group size in the live streaming channel, we first examine the average treatment effect of the Raid with Model (1). The result presented in Table 3 shows that existing viewers are more likely to be dormant ($\beta = -0.085$, $p < 0.01$) and post fewer comments ($\beta = -0.058$, $p < 0.01$). We observe that the number of existing commenters, on average, decreases by 8.1% (i.e., $\exp(-0.085) - 1$) after a Raid, and the comments per commenter for retained viewers decrease by about 5.6% (i.e., $\exp(-0.058) - 1$). Accordingly, we observe a negative average effect of group size on user engagement.

The result above reveals that the active existing viewers are less engaged, directly impacted by the increase in group size. Different from findings from asynchronous settings studied in seminal works, that is, Zhang and Zhu (2011), our results show that in synchronous settings such as live streaming, a larger group size could be detrimental to the engagement of active existing viewers, indicating a clear and dominant negative externality: congestion effect.

Table 3. Estimation Results for Main Effect

	Model (1)	
	ln(*CPC*)	ln(*CommCount*)
Raided	−0.058*** (0.007)	−0.085*** (0.015)
Constant	0.417*** (0.003)	0.873*** (0.006)
Playback fixed effects	Yes	Yes
Day of weeks	Yes	Yes
Time of the day	Yes	Yes
Session dummy	Yes	Yes
Number of playbacks	7,074	7,074
Observations	141,665	141,665
R^2	0.560	0.759

Notes. Robust standard errors in parentheses, clustered on the channel level. Note that we reran all applicable regressions to cluster the standard errors at the playback level, and the findings are consistent. For both dependent variables, we apply log transformation (i.e., $\log(x + 1)$) to seek percentage interpretation and accommodate skewness, throughout all the analyses.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Whereas the above results estimate the effect at the average level, the effect may be more nuanced based on the magnitude of the group size increase. To further investigate this, we conducted a subsample analysis to distinguish between the effects of large and small Raider groups. By dividing the data set into two subsamples—one with Raider sizes greater than the sample mean (namely *Raider: High*) and one with Raider sizes below the sample mean (namely *Raider: Low*)—we aimed to uncover how the magnitude of incoming viewers affects engagement metrics differently.

As shown in Table 4, the subsample analysis reveals an interesting pattern in engagement based on Raider group size. For the *Raider: High* subsample, the coefficients for both engagement measures are negative and significant (coefficients −0.062, $p < 0.01$ for ln(CPC) and −0.158, $p < 0.01$ for ln(CommCount)). Meanwhile, the *Raider: Low* subsample shows less negative impacts, with coefficients of −0.055 ($p < 0.01$) for *ln(CPC)* and −0.040 ($p < 0.01$) for *ln(CommCount)*. The results suggest that larger Raider groups potentially amplify the congestion effect, leading to greater decreases in engagement. In contrast, smaller groups are less disruptive in terms of the number of existing viewers who keep posting.

Notably, when we compare the coefficients of *Raided* for *ln(CommCount)* as the dependent variable, the difference is statistically significant ($p < 0.01$), suggesting that the larger the inflow size, the fewer existing commenters tend to engage in the conversations. However, when we compare the coefficients for *ln(CPC)*, the difference is statistically insignificant ($p > 0.1$). This finding suggests the group size congestion effect largely operates on the extensive margin instead of the intensive margin of engagement level. This aligns with social facilitation (Bond and Titus 1983), such that individuals feel a

Table 4. Subsample Analysis Results

	(1) *Raider: High*		(2) *Raider: Low*	
	ln(*CPC*)	ln(*CommCount*)	ln(*CPC*)	ln(*CommCount*)
Raided	−0.062*** (0.008)	−0.158*** (0.018)	−0.055*** (0.008)	−0.040*** (0.015)
Constant	0.480*** (0.003)	1.120*** (0.006)	0.379*** (0.003)	0.709*** (0.005)
Playback fixed effects	Yes	Yes	Yes	Yes
Day of weeks	Yes	Yes	Yes	Yes
Time of the day	Yes	Yes	Yes	Yes
Session dummy	Yes	Yes	Yes	Yes
Number of playbacks	3,292	3,292	3,782	3,782
Observations	65,900	65,900	75,765	75,765
R^2	0.530	0.770	0.557	0.710

Note. Robust standard errors in parentheses, clustered on the channel level.
 *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

decreased propensity to comment when they perceive they are "one among many" and that effect does not intensify along with the increase in group size.

5.2. Mediation Effects

We report the results of the mediation analysis in Table 5. In the first step, the Raid treatment increases topic incoherence by 0.133 ($p < 0.01$, Model (2-1)) and polarity by 0.015 ($p < 0.01$, Model (2-2)). While controlling for the direct effect (for *CPC* as DV, coefficient before Raided is −0.082, with $p < 0.01$; for *CommCount* as DV, coefficient before Raided is −0.211, $p < 0.01$ in Model (2-3)), greater incoherence has significant negative effects on engagement: Not considering the interaction term, a one unit increase in incoherence reduces comments per commenter by 6.67% (i.e., $\exp(-0.069) - 1$) and commenter count by 12.63% (i.e., $\exp(-0.135) - 1$), whereas a one unit increase in polarity increases comments per commenter by 3.25% (i.e., $\exp(0.032) - 1$) and commenter count by 4.19% (i.e., $\exp(0.041) - 1$).

To quantify the indirect effect, we further estimate the NIE, NDE, and proportion mediated, following Li et al.

(2022). The results above suggest a partial mediation, with 10.30% of the total effect on comments per commenter mediated by incoherence and polarity, whereas 8.42% of the total effect on commenter count is mediated by these factors. To provide a deeper understanding of these indirect effects, we calculated their relative contributions (i.e., mediated proportions). For comments per commenter, incoherence accounts for approximately 9.90% of the total effect, whereas polarity accounts for 0.42%. Similarly, for commenter count, incoherence mediates about 8.04% of the total effect, with polarity mediating 0.24%. The results indicate that cognitive factors (represented by incoherence) have a stronger mediating role compared with emotional factors (represented by polarity) in both engagement metrics. Specifically, the indirect effect through incoherence is about 23.6 times stronger than the effect through polarity for comments per commenter and about 33.5 times stronger for commenter count. This difference in magnitude shows the primary importance of cognitive processing challenges in explaining the negative impact of increased group size on user engagement in a live

Table 5. Estimation Results for Mediation Effect

	(2-1) *Incoherence*	(2-2) *Polarity*	(2-3)	
			ln(*CPC*)	ln(*CommCount*)
Raided	0.133*** (0.036)	0.015*** (0.005)	−0.082*** (0.007)	−0.211*** (0.015)
Incoherence			−0.069*** (0.001)	−0.135*** (0.002)
Polarity			0.032*** (0.004)	0.041*** (0.009)
Raided × *Incoherence*	—	—	−0.003*** (0.001)	0.049*** (0.003)
Raided × *Polarity*	—	—	−0.058*** (0.001)	−0.077*** (0.012)
Constant	1.639*** (0.014)	0.194*** (0.002)	0.638*** (0.003)	1.324*** (0.007)
Playback fixed effects	Yes	Yes	Yes	Yes
Day of weeks	Yes	Yes	Yes	Yes
Time of the day	Yes	Yes	Yes	Yes
Session dummy	Yes	Yes	Yes	Yes
Number of playbacks	7,074	7,074	7,074	7,074
Observations	141,665	141,665	141,665	141,665
R^2	0.479	0.141	0.623	0.831

Note. Robust standard errors in parentheses, clustered on the channel level.
 *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

streaming context. Nevertheless, the emotional aspect, while less pronounced, still plays a small role in the overall effect, highlighting the multifaceted nature of user engagement dynamics in response to exogenous increases in group size.

Altogether, the results suggest that first, both topic incoherence and polarity significantly increase after group size increases. Second, both topic incoherence and sentiment polarity have significant negative indirect effect on comments per commenter and commenter count level (i.e., NIE for polarity on *ln(CommCount)*: -0.00054; NIE for incoherence on *ln(CommCount)*: -0.01145; NIE for polarity on *ln(CPC)*: -0.00039; NIE for incoherence on *ln(CPC)*: -0.00956. NDE for polarity on *ln(CommCount)*: -0.226; NDE for incoherence on *ln(CommCount)*: -0.131; NDE for polarity on *ln(CPC)*: -0.09325; NDE for incoherence on *ln(CPC)*: -0.087). The above results suggest partial mediations through the two effects. The detailed calculation process with a robustness check is reported in Online Appendix 4.

It is also worth noting that the indirect effect through incoherence is more pronounced than the effect through polarity, given the relative magnitude of the two measures. Specifically, a one-unit increase in polarity indicates a switch from a neutral sentiment to an absolute positive or negative sentiment due to the probability-based nature of the measure. In contrast, a one-unit increase in incoherence represents only a slight change in the degree of topic incoherence based on its Euclidean distance-based construct. When considering the magnitude coefficients, the indirect effect through incoherence is expected to be much larger than the effect through polarity.

5.3. Effects with the Presence of Channel Moderators

The analysis results with human and bot moderators are presented in Table 6. In all models, the effects of *Raided* are highly consistent with our results in the main and mediation analysis. Results in Model (3-1) show that both human and bot moderators are effective in alleviating the negative main effect on comments per commenter, with human moderators (0.033, $p < 0.01$) relatively more effective than bot moderators (0.014, $p < 0.01$). Furthermore, Models (3-3) and (3-4) show that both human and bot moderators play a crucial role in managing incoherence and polarity. Specifically, Model (3-3) indicates that, although bot moderators significantly reduce topic incoherence (-0.091, $p < 0.05$), human moderators do not show a significant effect (-0.013, $p > 0.1$). In Model (3-4), human moderators are shown to significantly alleviate the positive main effect on polarity (-0.013, $p < 0.05$), whereas bot moderators do not exhibit a significant effect on this metric (-0.008, $p > 0.1$). These results highlight that human and bot moderators have very distinct roles in managing viewer engagement and maintaining content coherence in live streaming channels, with human moderators being more effective in reducing comment polarity and bot moderators in reducing topic incoherence.

Next, to further understand the nuanced differences between human and bot moderators across different sizes of incoming Raider groups, we conducted a subsample analysis. Table 7 presents the results for high and low Raider groups. Notably, for *Raider: High* groups, the presence of bot moderators significantly reduces topic incoherence ($\beta = -0.167$, $p < 0.05$) and comment polarity ($\beta = -0.016$, $p < 0.1$), whereas human moderators do not significantly affect these outcomes. In contrast, for *Raider: Low* groups, human moderators significantly alleviate the increase in polarity ($\beta = -0.016$, $p < 0.05$), but neither type of moderator significantly impacts topic incoherence. The findings suggest that moderators' efficacy is rather nuanced: Bot moderators are primarily effective in managing larger influxes of viewers by reducing topic incoherence, and to some extent, comment polarity, whereas human moderators are more effective at reducing comment polarity, when

Table 6. Estimation Results for Moderator Presence

	(3-1) *ln(CPC)*	(3-2) *ln(CommCount)*	(3-3) *Incoherence*	(3-4) *Polarity*
Raided	-0.137^{***} (0.012)	-0.187^{***} (0.025)	0.167^{***} (0.064)	0.028^{***} (0.008)
Raided × *Human*	0.033^{***} (0.007)	0.087^{**} (0.038)	-0.013 (0.049)	-0.013^{**} (0.006)
Raided × *Bot*	0.014^{**} (0.007)	0.007 (0.016)	-0.091^{**} (0.042)	-0.008 (0.005)
Constant	0.554^{***} (0.004)	1.130^{***} (0.008)	2.233^{***} (0.014)	0.258^{***} (0.002)
Playback fixed effects	Yes	Yes	Yes	Yes
Day of weeks	Yes	Yes	Yes	Yes
Time of the day	Yes	Yes	Yes	Yes
Session dummy	Yes	Yes	Yes	Yes
Number of playbacks	7,074	7,074	7,074	7,074
Observations	141,665	141,665	141,665	141,665
R^2	0.532	0.792	0.359	0.141

Note. Robust standard errors in parentheses, clustered on the channel level.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 7. Subsample Analysis Results

	Raider: High		Raider: Low	
	Incoherence	*Polarity*	*Incoherence*	*Polarity*
Raided	0.258** (0.125)	0.029*** (0.008)	0.049** (0.024)	0.025*** (0.009)
Raided × Human	0.048 (0.110)	−0.004 (0.009)	−0.051 (0.054)	−0.016** (0.007)
Raided × Bot	−0.167** (0.070)	−0.016* (0.009)	−0.009 (0.050)	−0.005 (0.006)
Constant	2.411*** (0.016)	0.259*** (0.002)	1.942*** (0.016)	0.255*** (0.002)
Playback fixed effects	Yes	Yes	Yes	Yes
Day of weeks	Yes	Yes	Yes	Yes
Time of the day	Yes	Yes	Yes	Yes
Session dummy	Yes	Yes	Yes	Yes
Number of playbacks	3,292	3,292	3,782	3,782
Observations	65,900	65,900	75,765	75,765
R^2	0.427	0.147	0.387	0.133

Note. Robust standard errors in parentheses, clustered on the channel level.
  *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

the incoming group is small. This finding further suggests bot moderator being a scalability solution for the congestion issue here.

6. Discussion and Conclusion

The massive user engagement in online synchronous content platforms, particularly for live streaming, brings both opportunities and challenges. As the streaming platforms and channels grow, viewers may be overloaded by the flood of information, and the live chat sessions may experience congestion. Thus, it is critical to understand how the group size of viewers would affect viewer engagement on live streaming platforms, which provides important implications for the design of platform features and interfaces.

Leveraging a panel data set collected from Twitch and the unique Raid functionality, we employ a difference-in-differences identification strategy to investigate the impact of group size on existing commenters' engagement. We observe that an influx of new viewers often prompts existing participants to reduce their commenting or exit the discussion, highlighting potential information congestion and overload (Schick et al. 1990, Simpson and Prusak 1995, Chen et al. 2009). Our mediation analysis reveals that this effect is driven by rising topic incoherence and comment polarity as the group grows. Moreover, although both bot and human moderators can moderate these adverse effects, bot moderators are particularly adept at curbing incoherence when the incoming group is large, whereas human moderators more effectively manage surges in comment polarity with smaller incoming groups.

Our findings advance existing literature on the impact of group size on asynchronous social media platforms by extending this stream of research to synchronous platforms. Whereas existing research on group size largely concludes that group size increases user engagement on asynchronous user-generated content platforms (Zhang and Zhu 2011, Qiu et al. 2016, De Reuver et al. 2018), our findings show that group size negatively affects existing user engagement in synchronous settings. We show suggestive evidence for the potential underlying mechanisms. The increase in group size increases the topic incoherence to the chat and, therefore, interrupts real-time communication, similar to what has been observed in the system interruption setting (Adamczyk and Bailey 2004). This scenario can be reasonably generalized to other live communication contexts, such as virtual conferences. Individual conference participants will raise fewer questions if they believe their questions may not be answered after an evaluation of the number of questions. In addition, we show that the increase in group size also raises the polarity of comments, which in turn leads to a decrease in user engagement. The heightened emotional reactions, both positive and negative, make it challenging to maintain coherent and meaningful interactions. This result extends some recent studies on social media platforms like Reddit (Xia et al. 2020) and Twitter (Salehabadi et al. 2022), which found that increased toxicity (i.e., extreme negative sentiment) can lead to higher engagement. Our results, showing negative mediation effects of both topic incoherence and comment polarity, highlight the unique dynamics of live streaming platforms and the importance of maintaining coherent and balanced discussions to foster user engagement as group sizes increase. Therefore, platforms and streamers need to pay careful attention to comment moderation to manage this dynamic and mitigate the negative effects of increased group size on engagement.

Further, we find that both bot and human moderators are effective in managing live streaming chatrooms, with varying efficacy. Live streaming channels that adopt human moderators can most effectively manage the polarity of comments, especially when the incoming group is small. Conversely, bot moderators are more effective in managing topic incoherence, helping to

maintain a coherent conversation flow despite the increase in group size. On Twitch, human moderators voluntarily enforce rules and reinforce social norms based on their judgment. Although they may be less efficient than bots in handling large volumes of comments, their nuanced understanding of the streaming context allows them to address polarizing comments more effectively. Bots, on the other hand, excel at mitigating incoherent comments, especially when the incoming group size is large, potentially due to their algorithmic and scalable nature. Our findings contribute to the recent discussion on the role of human and bot moderators in the literature (Ruckenstein and Turunen 2020, He et al. 2024). Although we do not conclude which type of moderator is overall more effective, our results provide valuable insights for platforms when they design features and deploy resources for content moderation.

Our research suggests several potential implications for live streaming platforms and possibly for other synchronous online environments with similar dynamics. The findings indicate that increases in group size, such as those caused by the Raid events, can have nuanced effects on user engagement. Platform designers and streamers might consider developing strategies to maintain engagement with existing commenters during periods of rapid audience growth. For instance, they could explore targeted interaction techniques or special events to avoid viewer loss. Our results also suggest the importance of effective content moderation in managing challenges arising from group size increases, particularly in maintaining chat room's topic coherence and managing comments' emotional polarity. The observed differences in effectiveness between human and bot moderators might inform platform decisions about moderation strategies. Although our findings are most directly applicable to live streaming contexts, they may offer insights for other real-time communication platforms, although further research would be needed to confirm broader applicability. Overall, these implications could potentially inform approaches to enhancing user experience and content governance in synchronous online group environments.

Our research has several limitations, which also open opportunities for future research. First, given the observational nature of this research, although we carefully designed the difference-in-differences study and implemented a series of robustness checks, we acknowledge that given the lack of a randomized experiment, we are unable to control for all confounders. For example, we might not be able to fully account streamers' decisions to Raid certain others. Second, given data limitation, the identities of Raiders are not available to us. This may be a great opportunity for a future field study wherein researchers could work with platforms to track Raiding viewers' behaviors after the Raids. Third, we focus on

one key type of interaction on the live streaming platform: commenting. Although we believe that commenting is critical in live streaming, future research may examine other activities, such as gifting and interaction with advertisements if such data are available. Fourth, our study specifically examines the effects of group size increases due to the nature of the Raid feature. We acknowledge that, although Raids offer a valuable opportunity to study group size effects, they have specific characteristics that may not apply to all group size increase scenarios. Raids represent an unexpected influx of viewers, which may have different implications compared with gradual audience growth. Future research could explore scenarios where group size decreases or fluctuates bidirectionally to further understand group size dynamics in other contexts. Finally, our mediation analysis reveals a partial mediation, suggesting that there could be other mechanisms not considered in this study. Similarly, the effect of the moderators in this study is only grounded on the proposed engagement measures. Future research may investigate additional factors to fully understand the underlying processes affecting viewer engagement.

Endnotes

[1] See https://www.dexerto.com/entertainment/twitch-streamer-permanently-banned-for-not-moderating-their-viewers-1506219/.

[2] We assess streamers' awareness of the Raid event and perform robustness checks to validate the exogenous setting. As reported in Appendix 3, the results confirm the validity of our empirical setting.

References

Adamczyk PD, Bailey BP (2004) If not now, when?: The effects of interruption at different moments within task execution. *Human Factors in Computing Systems* (Association for Computing Machinery, New York), 271–278.

Arguello J, Butler BS, Joyce E, Kraut RE, Ling KS, Rosé CP, Wang X (2006) Talk to me: Foundations for successful individual-group interactions in online communities. *Human Factors in Computing Systems* (Association for Computing Machinery, New York), 959–968.

Baek J, Shore J (2020) Forum size and content contribution per person: A field experiment. *Management Sci.* 66(12):5906–5924.

Bai X, Marsden JR, Ross WT, Wang G (2020) A note on the impact of daily deals on local retailers' online reputation: Mediation effects of the consumer experience. *Inform. Systems Res.* 31(4):1037–1492.

Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personality Soc. Psych.* 51(6): 1173–1182.

Blair A (2011) Information overload's 2,300-year-old history. Retrieved March 1, 2025, https://hbr.org/2011/03/information-overloads-2300-yea.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learn. Res.* 3:993–1022.

Bond CF, Titus LJ (1983) Social facilitation: A meta-analysis of 241 studies. *Psychol. Bull.* 94(2):265–292.

Cai J, Wohn DY (2022) Coordination and collaboration: How do volunteer moderators work as a team in live streaming

communities? *Proc. Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York).

Chandrasekharan E, Samory M, Jhaver S, Charvat H, Bruckman A, Lampe C, Eisenstein J, et al. (2018) The Internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proc. ACM Human-Computer Interaction* (Association for Computing Machinery, New York), 1–25.

Chen CP, Lin YC (2018) What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement. *Telematics Informatics* 35(1):293–303.

Chen D, Peterson JC, Griffiths TL (2017) Evaluating vector-space models of analogy. Preprint, submitted May 12, https://arxiv.org/abs/1705.04416.

Chen YC, Shang RA, Kao CY (2009) The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electronic Commerce Res. Appl.* 8(1):48–58.

Cheng J, Bernstein M, Danescu-Niculescu-Mizil C, Leskovec J (2017) Anyone can become a troll: Causes of trolling behavior in online discussions. *Proc. ACM Conf. Comput. Supported Cooperative Work* (Association for Computing Machinery, New York), 1217–1230.

Chmiel A, Sienkiewicz J, Thelwall M, Paltoglou G, Buckley K, Kappas A, Hołyst JA (2011) Collective emotions online and their influence on community life. *PLoS One* 6(7):e22207.

Christanto R, Miaou SG, Lin TL (2011) Reducing perceptual discontinuity in peer-to-peer live video streaming based on frame loss visibility information. *Proc. Internat. Conf. Electrical Engrg. Informatics* (IEEE, Piscataway, NJ), 1–6.

De Reuver M, Sørensen C, Basole RC (2018) The digital platform: A research agenda. *J. Inform. Tech.* 33(2):124–135.

Dean B (2021) How many users does clubhouse have? 40+ clubhouse stats (2021). Accessed April 18, 2021, https://backlinko.com/clubhouse-users.

Dean B (2024) X (Twitter) statistics: How many people use X? (2024). Accessed July 9, 2024, https://backlinko.com/twitter-users.

Ethayarajh K, Duvenaud D, Hirst G (2018) Towards understanding linear word analogies. Preprint, submitted October 11, https://arxiv.org/abs/1810.04882.

Foltz PW, Kintsch W, Landauer TK (1998) The measurement of textual coherence with latent semantic analysis. *Discourse Processes* 25(2–3):285–307.

Ford C, Gardner D, Horgan LE, Liu C, Tsaasan AM, Nardi B, Rickman J (2017) Chat speed op pogchamp: Practices of coherence in massive twitch chat. *Proc. CHI Conf. Extended Abstracts Human Factors Comput. Systems* (Association for Computing Machinery, New York), 858–871.

Grosz BJ, Weinstein S, Joshi AK (1995) Centering: A framework for modeling the local coherence of discourse. *Comput. Linguist.* 21(2):203–225.

Guan Z, Hou F, Li B, Phang CW, Chong AYL (2022) What influences the purchase of virtual gifts in live streaming in China? A cultural context-sensitive model. *Inform. Systems J.* 32(3):653–689.

Guo C, Kim TH, Susarla A, Sambamurthy V (2020) Understanding content contribution behavior in a geosegmented mobile virtual community: The context of Waze. *Inform. Systems Res.* 31(4):1398–1420.

Hamilton WA, Garretson O, Kerne A (2014) Streaming on Twitch: Fostering participatory communities of play within live mixed media. *Proc. SIGCHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York), 1315–1324.

Hayes AF (2017) *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (Guilford Publications, New York).

He Q, Hong Y, Raghu TS (2024) Platform governance with algorithm-based content moderation: An empirical study on Reddit. *Inform. Systems Res.*, ePub ahead of print August 21, https://doi.org/10.1287/isre.2021.0036.

Hong Y, Peng J, Burtch G, Huang N (2021) Just DM me (politely): Direct messaging, politeness, and hiring outcomes in online labor markets. *Inform. Systems Res.* 32(3):786–800.

Hu Y, Talamadupula K, Kambhampati S (2013) Dude, srsly?: The Surprisingly formal nature of Twitter's language. *Proc. Internat. AAAI Conf. Web Social Media*, vol. 7, 244–253.

Huang N, Burtch G, He Y, Hong Y (2022) Managing congestion in a matching market via demand information disclosure. *Inform. Systems Res.* 33(4):1196–1220.

Jones Q, Ravid G, Rafaeli S (2004) Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Inform. Systems Res.* 15(2):194–211.

Katz ML, Shapiro C (1985) Network externalities, competition, and compatibility. *Amer. Econom. Rev.* 75(3):424–440.

Katz ML, Shapiro C (1994) Systems competition and network effects. *J. Econom. Perspective* 8(2):93–115.

Kaytoue M, Silva A, Cerf L, Meira W, Raïssi C (2012) Watch me playing, I am a professional. *Proc. 21st Internat. Conf. Companion World Wide Web* (Association for Computing Machinery, New York), 1181–1188.

Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci. USA* 111(24):8788–8790.

Kuang L, Huang N, Hong Y, Yan Z (2019) Spillover effects of financial incentives on non-incentivized user engagement: Evidence from an online knowledge exchange platform. *J. Management Inform. Systems* 36(1):289–320.

Kumar R, Mahdian M, McGlohon M (2010) Dynamics of conversations. *Proc. 16th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, 553–562.

Kusner MJ, Sun Y, Kolkin NI, Weinberger KQ (2015) From word embeddings to document distances. *Proc. 32nd Internat. Conf. Machine Learn.*, vol. 37 (PMLR, New York), 957–966.

Levy O, Goldberg Y (2014) Neural word embedding as implicit matrix factorization. *Adv. Neural Inform. Processing Systems* 27:2177–2185.

Li Y, Guo Y (2021) Virtual gifting and danmaku: What motivates people to interact in game live streaming? *Telematics Informatics* 62.

Li X, Grahl J, Hinz O (2022) How do recommender systems lead to consumer purchases? A causal mediation analysis of a field experiment. *Inform. Systems Res.* 33(2):620–637.

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, et al. (2019) RoBERTa: A robustly optimized BERT pretraining approach. Preprint, submitted July 26, https://arxiv.org/abs/1907.11692.

Lu S, Yao D, Chen X, Grewal R (2021) Do larger audiences generate greater revenues under pay what you want? Evidence from a live streaming platform. *Marketing Sci.* 40(5):964–984.

Lyu K, Kim H (2016) Sentiment analysis using word polarity of social media. *Wireless Personal Comm.* 89(3):941–958.

Matias JN (2019) The civic labor of volunteer moderators online. *Soc. Media Soc.* 5(2).

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, vol. 26 (Curran Associates, Inc., Red Hook, NY).

Mohan S, Guha A, Harris M, Popowich F, Schuster A, Priebe C (2017) The impact of toxic language on the health of Reddit communities. Mouhoub M, Langlais P, eds. *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, vol. 10233 (Springer, Cham, Switzerland), 51–56.

Muhammad A, Wiratunga N, Lothian R (2016) Contextual sentiment analysis for social media genres. *Knowledge-Based Systems* 108:92–101.

Pennington J, Socher R, Manning CD (2014) GloVe: Global vectors for word representation. *Proc. Conf. Empirical Methods Natural Language Processing* (Association for Computational Linguistics), 1532–1543.

Pessoa L (2008) On the relationship between emotion and cognition. *Nature Rev. Neurosci.* 9(2):148–158.

Qiu L, Benbasat I (2005) Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *Internat. J. Human-Comput. Interaction* 19(1):75–94.

Qiu L, Tang Q, Whinston AB (2016) Two formulas for success in social media: Learning and network effects. *J. Management Inform. Systems* 32(4):78–108.

Ruckenstein M, Turunen LLM (2020) Re-humanizing the platform: Content moderators and the logic of care. *New Media Soc.* 22(6):1026–1042.

Salehabadi N, Groggel A, Singhal M, Roy SS, Nilizadeh S (2022) User engagement and the toxicity of tweets. Preprint, submitted November 7, https://arxiv.org/abs/2211.03856.

Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Inform. Processing Management* 24(5):513–523.

Schick AG, Gordon LA, Haka S (1990) Information overload: A temporal approach. *Accounting Organ. Soc.* 15(3):199–220.

Seering J, Kairam SR (2024) Who moderates on Twitch and what do they do? Quantifying practices in community moderation on Twitch. *Proc. ACM Human-Comput. Interactions* (Association for Computing Machinery, New York), 7.

Seering J, Flores JP, Savage S, Hammer J (2018) The social roles of bots: Situating bots in discussions in online communities. *Proc. ACM Human-Comput. Interactions* (Association for Computing Machinery, New York), 1–29.

Seering J, Wang T, Yoon J, Kaufman G (2019) Moderator engagement and community development in the age of algorithms. *New Media Soc.* 21(7):1417–1443.

Shewale R (2024) Twitch statistics in 2024 (users, revenue & streamers). Accessed July 9, 2024, https://www.demandsage.com/twitch-users/.

Simpson CW, Prusak L (1995) Troubles with information overload: Moving from quantity to quality in information provision. *Internat. J. Inform. Management* 15(6):413–425.

Singha J, Das K (2013) Indian Sign Language recognition using Eigen value weighted Euclidean distance based classification technique. *Internat. J. Adv. Comput. Sci. Appl.* 4(2).

Sjöblom M, Hamari J (2016) Why do people watch others play video games? An empirical study on the motivations of Twitch users. *Comput. Human Behav.* 75:985–996.

Speier C, Valacich JS, Vessey I (1999) The influence of task interruption on individual decision making: An information overload perspective. *Decision Sci.* 30(2):337–360.

Srinivasan KB, Danescu-Niculescu-Mizil C, Lee L, Tan C (2019) Content removal as a moderation strategy: Compliance and other outcomes in the changemyview community. *Proc. ACM Human-Comput. Interactions* (Association for Computing Machinery, New York), 3.

Stephen B (2020) The lockdown live-streaming numbers are out, and they're huge. Accessed October 10, 2020, https://www.theverge.com/2020/5/13/21257227/coronavirus-streamelements-arsenalgg-twitch-youtube-livestream-numbers.

Tan X, Wang Y, Tan Y (2019) Impact of live chat on purchase in electronic markets: The moderating role of information cues. *Inform. Systems Res.* 30(4):1248–1271.

Toubia O, Stephen AT (2013) Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter? *Marketing Sci.* 32(3):368–392.

TwitchTracker (2024) Twitch watch time statistics. Accessed July 9, 2024, https://twitchtracker.com/statistics/watch-time.

Valeri L, VanderWeele TJ (2013) Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psych. Methods* 18(2):137–150.

VanderWeele TJ (2016) Mediation analysis: A practitioner's guide. *Annu. Rev. Public Health* 37:17–32.

Witte SP, Faigley L (1981) Coherence, cohesion, and writing quality. *College Composition Comm.* 32(2):189–204.

Wohn DY (2019) Volunteer moderators in Twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience. *Proc. CHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York), 1–13.

Wulf T, Schneider FM, Beckert S (2018) Watching players: An exploration of media enjoyment on Twitch. *Games Culture* 15(3):328–346.

Xia Y, Zhu H, Lu T, Zhang P, Gu N (2020) Exploring antecedents and consequences of toxicity in online discussions: A case study on Reddit. *Proc. ACM Human-Comput. Interactions* (Association for Computing Machinery, New York), 4.

Yang M, Ren Y, Adomavicius G (2019) Understanding user-generated content and customer engagement on Facebook business pages. *Inform. Systems Res.* 30(3):839–855.

Zhang XM, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.

Zhang H, Yang Y, Zhao J (2023) Does game-irrelevant chatting stimulate high-value gifting in live streaming? A session-level perspective. *Comput. Human Behav.* 138.

Zhao K, Hu Y, Hong Y, Westland JC (2021) Understanding characteristics of popular streamers on live streaming platforms: Evidence from Twitch.tv. *J. Assoc. Inform. Systems* 22(4):1076–1098.

Zhao K, Lu Y, Hu Y, Hong Y (2023) Direct and indirect spillovers from content providers' switching: Evidence from online livestreaming. *Inform. Systems Res.* 34(3):847–866.