Computers in Human Behavior Reports 13 (2024) 100358

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Effects of negativity type and active involvement on the likelihood of responding to negativity in live stream chats

Teodora M. Mihailova [a,*], Jeffrey A. Hall [b]

[a] School of Humanities and Social Sciences, Emporia State University, United States
[b] Department of Communication Studies, University of Kansas, United States

ARTICLE INFO

Keywords:
Active
Passive involvement
Ambiguous negativity
Calling out
Experiment
Live chats
SIDE model

ABSTRACT

This study explores whether chat negativity and the degree to which live chat rule sets encourage active (vs. passive) involvement influence participants' willingness to react to negative behavior within video game live streams. Using the Social Identity Model of Deindividuation Effects (SIDE) and an experimental design, this study examines chat participants' likelihood of calling out and reporting negative behaviors. A 2x3 experimental design manipulated type of negativity (i.e., clear/ambiguous) and framing of community-specific rules of users' role in responding to norm violations (i.e., active involvement/passive involvement/control). Results suggest clear negativity was associated with a higher likelihood of calling out/reporting. Active involvement interacted with degree of negativity: when live chat rule sets encouraged active (vs. passive) involvement, participants were more likely to call out clear negativity and less likely to call out ambiguous negativity. Furthermore, there was support for the hypothesis that social identification moderated the relationship between type of negativity and likelihood of response, whereby participants with higher social identification were more likely to respond to clear negativity and less likely to respond to ambiguous negativity. Finally, participants' perceptions of group norms in the hypothetical communities were affected by prior experience and chat activeness, but not by type of negativity or active (vs. passive) involvement.

1. Introduction

Negative online behavior (e.g., flaming, impoliteness, incivility) has been a concern since the beginning of computer-mediated communication (Hardaker, 2010). Such behavior has been studied across various contexts and platforms, including online communities and forums (Bergstrom, 2011; Hardaker, 2010; Suh & Wagner, 2013; Teneketzi, 2021), online gaming (Cook, Conijn, Schaafsma, & Antheunis, 2019), and live-streaming platforms (Seering, Kraut, & Dabbish, 2017). Live streaming has been studied across the world, including Taiwan (Chen & Lin, 2018), China (Hu, Zhang, & Wang, 2017), South Korea (Yu, Jung, Kim, & Jung, 2018), and the U.S., and on platforms including YouTube live (Guarriello, 2019) and Twitch.tv (Taylor, 2018).

With the prevalence of and concern about negative online behavior, research has sought ways to curb it and encourage positive interactions. Community moderators are an important part of shaping positive interaction, but Seering, Wang, Yoon, and Kaufman (2019) question the ethics of building positive communities by relying solely on the volunteer labor of moderators, described by some as a "second job" (p. 1427). Furthermore, moderation is a reactive (rather than a proactive) approach to community management and considering the documented issues of scalability and ethics within moderation (Seering et al., 2019), it is, by itself, insufficient. Thus, it is important to explore what factors might encourage ordinary users to respond to chat negativity by intervening or calling out problematic behavior. This study extends research on online negativity and community moderation by employing a 3x2 experimental design to explore whether Twitch.tv users can be motivated to act in response to negative online behavior by published chat rules.

This study contributes to theory by extending the Social Identity Model of Deindividuation Effects (SIDE) (Postmes, Spears, & Lea, 1998) to live streaming and live chats. Negative comments made on websites and forums (e.g., Chung, 2019; Rösner & Krämer, 2016) might later be subjected to moderation, called out by others, or otherwise punished. With the synchronous nature of live streaming, if moderation or calling out does not occur in a timely manner, the interaction fades away unchallenged, which might impact users' experiences and perceptions of group norms (e.g., London, Crundwell, Eastley, Santiago, & Jenkins,

* Corresponding author. Emporia State University, Emporia, KS, 66801, United States.
E-mail address: tmihailo@emporia.edu (T.M. Mihailova).

https://doi.org/10.1016/j.chbr.2023.100358
Received 19 July 2023; Received in revised form 20 October 2023; Accepted 1 December 2023
Available online 5 December 2023

2019). Indeed, post factum moderation is largely invisible to users. During live streaming, it is crucially important for the moderation response to be provided in the moments following the interaction if the goal is to foster a more positive environment. Past research, however, has not explored whether community engagement rules can harness users' sense of community and social identity toward positive intervention. This study examined whether published rules by content creators on Twitch.tv, who are the center of their respective communities, can motivate users through social identification with the streaming community.

2. SIDE model and online negativity

Despite the long tradition of studying online negativity, its definition and categorization is inconsistent and fraught (Cook, Schaafsma, & Antheunis, 2018; Cruz, Seo, & Rex, 2018). For instance, when looking at impoliteness and incivility, researchers differ on whether they consider them separately or as a single issue, as well as what the normative stance on appropriateness should be (Masullo Chen, Muddiman, Wilner, Pariser, & Stroud, 2019). Some draw the line at any profanity and impoliteness while others only consider hateful or undemocratic speech unacceptable. Masullo Chen et al. (2019) argue that focusing on uncivil language and assuming it can be objectively measured misses the point: we do not want to reduce incivility because it is bad in and of itself, but because some of its manifestations might cause harm by proliferating hate and social exclusion, particularly for individuals from marginalized groups. Instead, researchers recommend understanding and training people to act against the harmful outcomes of online incivility, as well as empowering communities to decide for themselves what should be acceptable. Online spaces and communities, such as those centered around gaming, are examples of spaces where negativity can include both milder incivility such as cursing and writing in capital letters and aforementioned harmful and exclusionary content; thus, they stand to benefit from Masullo Chen et al.'s (2019) suggestions. While the development of advanced community moderation practices to manage online negativity is practical and important (Gillepsie, 2017), research should explore other ways to empower communities by encouraging users to be proactive.

Before counteracting negativity online, we should understand where it comes from. The SIDE model is pertinent and has shown considerable empirical support (e.g., Postmes, Spears, Sakhel, & de Groot, 2001; Rösner & Krämer, 2016). Social identity refers to the degree to which community members identify as members of a group versus identifying as individuals (Postmes et al., 2001) and attach value to that membership (Suh & Wagner, 2013). In mediated interaction where participants do not have all available cues that they would in a face-to-face interaction, this social identity becomes more salient. This is associated with compliance with in-group norms, which can result in negative behavior when local norms support such behavior (Postmes et al., 1998, 2001). The SIDE model has previously been applied to interaction and norm adoption in online comment sections (Chung, 2019; Rösner & Krämer, 2016). Seering, Ng, Yao, and Kaufman (2018) identify SIDE as an area of emergent scholarship in computer-mediated communication in the areas of community and collaborative digital contexts; they also pose a call for further applications of SIDE to such contexts, which the present work seeks to answer. SIDE is also relevant to Twitch.tv, where streamers' fans are anonymous members of their online communities (Blight, 2016), each with distinctive norms (Seering et al., 2017).

Knowing when calling out potentially negative behavior is required is no less important than understanding what counts as inappropriate behavior. When considering negative behaviors, SIDE is sensitive to specific community norms and contexts. Prior research has applied various terms for different negative behaviors, including flaming, trolling, spamming, and cyberbullying (Hardaker, 2010), but a clear typology has not emerged (Cook et al., 2018; Cruz et al., 2018). It is challenging to determine whether and under what conditions these behaviors are truly negative. Cruz et al. (2018) "challenge the dominant view of trolling as merely a value-destroying anti-social practice" (p. 24) in favor of acknowledging that such transgressive and provocative behaviors can be legitimate mechanisms of engagement. Karhulahti (2016) demonstrated that on Twitch.tv behaviors we typically see as negative, such as pranks and trolling, could be normative. Seering et al. (2017) suggested that certain live chat practices, such as spamming emoticons repeatedly, might be prohibited in some communities but encouraged in others. On Twitch.tv, Taylor (2018) suggests this behavior is a form of cheering: "a fast-scrolling chat window, filled with text that wasn't conversational but full of excited exclamations, repetitive emoticons, and memes, could be seen as akin to the cheering one would find in a sports stadium" (p. 42). Therefore, what constitutes negativity is not always clear-cut and should be considered at the community level (Masullo Chen et al., 2019). Thus, it is important for researchers to explore which types of negativity require intervention in investigations of how to encourage community members to intervene proactively. In their discussion of potential future applications of social identity approaches to communal and collaborative online contexts, Seering et al. hypothesize through SIDE that toxic behavior might be mitigated by making salient the less toxic elements of the social identity of the group (2018, p. 201–17). We hope to explore a similar relationship where we test rules about negative behavior and an encouragement of proactive involvement as positive elements of a hypothetical social identity to determine whether they can motivate users to step in when witnessing negativity (and thus in turn mitigate toxic behavior without increasing moderator burden).

3. Twitch.tv

A prominent gaming-oriented live-streaming platform, Twitch.tv is at the intersection of discussions of negative behaviors within online gaming and live-streaming (Taylor, 2018). Twitch.tv is a popular medium that grown since the outbreak of the COVID-19 pandemic (Narassiguin & Garnès, 2020). In it streamers broadcast live videos of themselves, usually playing video games, alongside footage from within the game. Live chat interaction can occur near-synchronously as viewers can comment, talk to each other, and address the streamer as the video unfolds. Viewers have varied motivations for using the platform (e.g., seeking information; Sjöblom & Hamari, 2016), yet Hilvert-Bruce, Neill, Sjöblom, and Hamari (2018) showed that Twitch.tv viewers were more socially motivated than consumers of traditional mass media. Blight (2016) found that viewers form online communities surrounding their favorite streamers, as in other live streaming platforms (Guarriello, 2019; Yu et al., 2018).

3.1. Twitch.tv as community

These social motivations and the formation of communities suggest that social identification processes might be particularly pertinent to Twitch.tv. For instance, Diwanji et al. used social identity theory in their exploration of information behavior and copresence in Twitch,tv live chats, arguing that social identity processes are at play when it comes to membership in those communities because to many participants "the stream is not only a source of entertainment, but a digital community to which they belong and regularly contribute" (2020, p. 6). Diwanji et al.'s (2020) qualitative findings showing how participants perform in-group norms through their use of emotes specific to their communities support this idea. Additionally, Wohn and Freeman (2020) showed that streamers who identify with particular social groups (e.g., LGBTQ streamers, streamers of color) expected their audience to be of the same group and advertised themselves to them accordingly. Thus, Twitch.tv communities might share multiple overlapping social identities – not only a shared identity as members of the community, but also potentially other shared identities in real life. Online communities can have specific norms, rules (Seering et al., 2017) and language (Marvin,

1995), so they have context-specific approaches to moderation and norms about what constitutes negativity.

3.2. Norms and moderation on Twitch.tv

There are two mechanisms through which rules and norms affect how people behave and understand what is acceptable: *injunctive norms* (i.e., what ought to be; what is supported or seen as unacceptable, enforced through moderation or rules) and *descriptive norms* (i.e., what is; what is visibly practiced by community members) (Cialdini, Reno, & Kallgren, 1990). Both appear to influence how people behave in live chats. Moderation and example-setting can be effective tools in shaping positive interactions online. Suh and Wagner (2013) show that users tasked with reviewing post content, warning rule-breakers and discontinuing memberships might curb negative behaviors in some communities (injunctive norms). Rösner and Krämer's (2016) experiment showed that verbal aggressiveness in preceding comments predicted verbal aggressiveness in subsequent comments (descriptive norms).

Taylor (2018) noted that Twitch.tv moderators can act proactively or reactively to shape live chat interactions. Seering et al. (2017) showed that more positive interactions in Twitch.tv live chats can be shaped via moderation and example-setting: in the short-term, proactive and reactive moderation tools effectively deterred negative behavior, whereas positive behavior modeled by users, particularly ones of higher authority, fostered more positive subsequent interactions through imitation. Conversely, London et al. (2019) show that when harassment is not moderated consistently, it continues to occur. Consistently with the SIDE model, these are examples of creating norms by moderating interactions (injunctive norms) and making practicing prosocial behavior more common (descriptive norms). Because Twitch.tv affords chat users anonymity and houses online communities where social identity can be made salient, that social identity can then influence users to act more consistently with those norms.

As online communities grow, small-scale moderation practices are insufficient (Gillepsie, 2017). In larger communities, platforms employ outsourced and automated moderation (Jhaver, Birman, Gilbert, & Bruckman, 2019), which scale to larger communities, but often lack sophistication in interpreting negativity correctly in context (Rosenthal & Belmas, 2021). While outsourcing or automating moderation have become commonplace, they miss some important "social nuances of moderation" (Seering et al., 2019, p. 1434). The nuance offered by moderation from community insiders who understand the norms and practices specific to the community are very valuable. Thus, proactive involvement of regular users who are not moderators, such as positive example-setting (Seering et al., 2017), can be effective and have the potential to shape online interactions while reducing moderator burden.

Regular community members can choose to respond to live chat negativity publicly, but they might often be reluctant to step in (London et al., 2019), choose to report it privately (Seering et al., 2019), or choose not to respond (Mihailova, 2022; Rebollo-Catalan & Mayor-Buzon, 2020). Volunteer community moderators in Seering et al.'s (2019) interviews noted that community members were more likely to flag content for moderation (which is only visible to the moderator), than to call people out (which involves interacting directly and is visible to chat participants and viewers). Regular live chat participants on Twitch.tv rarely get involved when witnessing hostility (Mihailova, 2022). Indeed, over a third of adolescents who observed cyber violence against girls or women self-reported doing nothing (Rebollo-Catalan & Mayor-Buzon, 2020). This points to the broader importance of empowering communities to regulate themselves. People might be reluctant to respond to inappropriate behavior, especially publicly, but as demonstrated by London et al. (2019), failing to moderate in cases of harassment in the moment means that such behaviors will continue to pop up again, seemingly without punishment. One example in London et al.'s (2019) study found after a community member called for a moderator to penalize someone, others join them in agreement. Thus, the type of involvement to be encouraged in regular community members in such situations should be responding to norm-violating negativity in a timely manner, such as by calling out the person in the public chat directly, or else calling for moderator response.

3.3. Rules as an avenue to empower regular Twitch.tv users to act

In this study, normative social influence is increased by anonymity under conditions of deindividuation when social identity is made salient (Postmes et al., 2001), but not when it is not made salient (Perfumi, Bagnoli, Caudek, & Guazzini, 2019). Without indication that social identity is salient to participants, a strong normative influence effect cannot be expected, which is a limitation of this approach. Prior experimental studies have applied the SIDE model to zero-history groups (i.e., by simulating community conditions without activating a preexisting group identity) by measuring identification (e.g., Chung, 2019; Rösner & Krämer, 2016) - even with research-generated segments of user contribution (e.g., Rösner & Krämer, 2016). Since the current study seeks to apply the SIDE model by simulating group membership in a zero-history condition, it will follow a similar approach to those set by these researchers (Chung, 2019; Rösner & Krämer, 2016). Similarly to Rösner and Krämer (2016) and Chung (2019), this study will take preliminary measures of experience with the platform in question (which in this case will pertain to the Twitch.tv platform) and will use this measure of perceived social identification as a proxy for social identity.

Specific to Twitch.tv, live chat negativity might be shaped by published chat rules. Twitch.tv streamers can shape community involvement injunctively by publishing rules on their channel, instructing how to participate in live chats. Rule sets should be studied because they constitute injunctive norms and frame users' understanding of the community's rules and norms (Mihailova, 2022). Particularly on a synchronous medium such as Twitch.tv, unchallenged negativity (i.e., without reprimand /injunctive norm enforcement) might set a negative example via shaping viewers' perceptions of norms (London et al., 2019). Channel-specific rules provided by streamers reflect platform-level rules but are also customized to the needs of the specific community. In their exploration of the channel and chat rules of Twitch.tv channel communities, Cai and Wohn (2019, November) showed that the transparency of the rules and the frequency with which they are communicated affect the conduct of viewers of a channel. They also state that "the rules that are encouraged by the streamer may dictate the values of the micro-communities" (Cai & Wohn, 2019, November, p. 293). Rules can be an important part of community identity. For instance, Myles, Benoit-Barné, and Millerand (2020) showed how members of one Reddit community structured their posts in such a way that performatively demonstrated their commitment to the community by making it obvious that they are adhering to its rules and values. Streamers shape community norms through published rules, and when participants' social identity is tied to membership, they may be more likely to adhere to these rules and norms. Thus, published rules and social identity likely both influence subsequent action.

4. The present investigation

This investigation seeks to determine whether channel rules specifying more active (as opposed to passive) involvement might be able to encourage regular participants to step in when witnessing rule-violating live chat negativity. This is an important avenue for exploration because increasing the likelihood for ordinary users to step in when witnessing negativity has the potential to reduce moderator burden while still providing a context-sensitive and timely response. This study is theoretically grounded in the SIDE model (Postmes et al., 1998, 2001) because this approach demonstrates the way social identity can strengthen compliance and rule enforcement in anonymous online interactions such as those in stream live chats. This investigation furthers SIDE research by applying it to the synchronous medium of stream live

chats, as well as by using rules as an instantiation of injunctive norms and a potential source of community identification.

This experiment examines live chat participants' likelihood of taking action (i.e., reporting the behavior to the streamer or moderators or confronting the person in chat directly) in response negative behaviors (e.g., curse words, exclusionary language) by fellow participants as dictated by the degree of passive or active involvement encouraged by community rules. Online behaviors range considerably in degree of antisociality, and users' perceptions of their inappropriateness also vary. As discussed above, not all negativity is harmful and different online communities have different criteria for what should be considered inappropriate. Thus, clearly negative behaviors which violate both generic ideas of civility and the community-specific norms (as articulated in the manipulated rules) are more easily identifiable as inappropriate than ambiguous remarks that might be considered uncivil in some contexts but do not violate the community-specific norms. We vary type of negativity (i.e., ambiguous vs. clearly hostile) and predict:

H1. Type of negativity will be positively associated with likelihood of taking action, whereby participants presented with a behavior which is clearly negative will be more likely to call it out when compared to participants presented with an ambiguous negative behavior.

We test whether participants were influenced by the way rule sets framed what their involvement should be in shaping the community when witnessing negativity. According to SIDE, social identity drives conforming to injunctive group norms. Rule sets instruct Twitch.tv users how to behave. As such, their content could be a factor that affects participants' decisions regarding whether to respond to negative behavior. Rules might ask participants to be more passive or ignore transgressors (passive involvement) or to be more proactive and call them out (active involvement). We conceptualize active and passive involvement as different stances that rules can encourage chat participants to adopt*: active involvement* implies a sense of responsibility and ownership of the community, and thus an expectation that a participant should be proactive in responding to norm violations and thus shaping the community by providing an example of enacted injunctive norms; *passive involvement* implies a role more akin to that of a spectator, whereby the conduct and norm violations of others are outside one's own concerns and thus require no response. McMillan and Chavis's (1986) sense of community construct includes an *influence* dimension which suggests that community members feel that they can affect and shape their community; being encouraged to be active in response to transgressions relates to this because it might enhance this sense of influence. The degree to which rules encourage participants to be actively involved in responding to transgressions was manipulated (i.e., active involvement vs. passive involvement vs. a control condition without rules about involvement). Rule sets that encourage more active involvement in participants should increase their willingness to uphold norms:

H2. Encouraging active involvement will be positively associated with likelihood of taking action, whereby participants presented with rules that encourage an active stance in shaping the community will be more likely to respond to negative behaviors when compared to participants presented with rules that encourage a passive stance.

As discussed above, the goal of encouraging community members to combat negativity proactively should not be undertaken without an understanding of when intervention is not required. The difference between the clear and ambiguous negativity conditions lies in whether the segment violates both community rules and general social norms (i.e., clear negativity) or only violates generic norms applicable to other contexts, but not the specific norms of the community (i.e., ambiguous negativity). A person unfamiliar with the specific community norms might consider ambiguous negativity inappropriate based on common knowledge, whereas someone who identifies with and understands the community and its norms would know that it is permissible. Because active involvement gives participants the responsibility of holding others accountable, it also tasks them with understanding the rules they should hold them accountable to; thus, we predict an interaction effect between the two main effects such that different combinations of clear/ambiguous negativity and active/passive involvement will have different effects on likelihood of taking action:

H3. Encouraging active involvement will interact with type of negativity in affecting likelihood of taking action, such that when the rules encourage active (as opposed to passive) involvement, there will be a greater difference in likelihood of taking action between the clear negativity versus the ambiguous negativity condition.

We predict that the social identification participants feel with the hypothetical communities will act as a moderating variable. Based on SIDE, rules encouraging active involvement in conjunction with higher group identification should give participants a stronger sense of ownership and urge to uphold norms (McMillan & Chavis, 1986; Perfumi et al., 2019). We follow prior research (e.g., Chung, 2019; Rösner & Krämer, 2016) in applying SIDE to a zero-history group by simulating membership and measuring social identification. Thus:

H4. The degree of social identification (measured within participants) will moderate the relationship between (a) the type of negativity and likelihood of taking action, and (b) encouraging active involvement and likelihood of taking action, whereby when social identification is high, participants' involvement will be higher in conditions of (a) clearer negativity and (b) rules encouraging active (as opposed to passive) involvement.

We also predict a three-way interaction (i.e., type of negativity, active/passive involvement, social identification) predicting likelihood of taking action:

H5. Likelihood of taking action will be highest when type of negativity is clear, rules encourage active (as opposed to passive) involvement, and social identification is high.

Finally, this investigation is interested in the effects of encouraging active involvement and negativity on perceived group norms. In addition to actual or intended action (i.e., intervening, calling out, etc.), another interesting outcome is the degree to which participants perceive the norms of the hypothetical community based on the rules and interactions presented to them, because such understanding is needed for meaningful proactive intervention appropriate to the community. Participants might gain an understanding of community norms by observing descriptive norms (i.e., seeing how community members behave, herein witnessing the segments presented to them) or injunctive norms (i.e., seeing evidence of how members are supposed to behave based on the rules presented to them). The interactions and rules are designed so the clear negativity condition can easily be recognized as against the rules, whereas the ambiguous negativity condition might be inappropriate in other more generic contexts but is not in clear violation of the rules. Type of negativity was predicted to affect this relationship because participants presented with rules and then asked to evaluate a behavior in the gray area might doubt their interpretation of the rules, whereas participants asked to do a straightforward evaluation of a clearly hostile behavior might be more confident that they understood the rules correctly:

H6. Type of negativity will be positively associated with perceived group norms, whereby participants presented with a behavior which is clearly negative will be more likely to perceive these behaviors as not allowed compared to participants presented with an ambiguous negative behavior.

Encouraging active involvement was predicted to affect this relationship because unlike participants who see rules asking them to be proactive, participants asked by the rules to be more passive might not care as much about how the rules should be interpreted and applied

because it is not their responsibility to enforce them:

H7. Encouraging active involvement will be positively associated with perceived group norms, whereby participants presented with rules encouraging active (as opposed to passive) involvement will be more likely to perceive negative behaviors as not being allowed in the community compared to participants presented with rules that encourage passive involvement.

5. Methods and procedures

Data was collected February through November 2022. Undergraduate student participants were recruited from communication courses at a midwestern university. Participants had to be familiar with video game live streaming sites and received extra credit (< .5% of final grade).

After a required screening question about live stream viewership and measures of prior viewership experience and live chat activeness, participants were exposed to a set of sample live chat rules featuring active involvement vs. passive involvement vs. control, followed by a researcher-generated chat interaction excerpt containing negativity (Rösner & Krämer, 2016). Live chat excerpts were designed to resemble a screenshot from a Twitch.tv live stream, with game footage, a streamer's face cam, and an accompanying live chat. The excerpt contained either ambiguous negativity or clearly hostile remarks. In the clear negativity condition, the negative behavior was explicitly prohibited by the rules and was not exhibited by any of the preceding rule-conforming chat messages. After reading the hypothetical scenario, participants completed manipulation check measures, likelihood of taking action, perceptions of group norms, how active they felt their involvement should be based on the rules, the extent to which they identified with chat participants who were *not* violating norms, and a demographic questionnaire. See supplemental materials.

Participants were flagged for instances of straight-lining (i.e., same responses on one of two scales that included negatively worded items) or falling outside the mean ±2.55 *SD* range of completion time. Participants were removed if flagged two or more times, failed the screener, or provided incomplete responses. The initial sample included 291 participants, and the final sample had 212.

5.1. Participants

Of the final sample, 11 participants (4.9%) did not complete demographic measures. Those who did were 46.2% female, 52.4% male, and 1.4% non-binary. Participants' mean age was 19.3 years old (*SD* = 1.87, range 18–33). Participants could select any number of racial/ethnic categories that applied to them: 75.8% selected White; 8.1% Hispanic/Latinx; 4.9% African-American/Black; 4.9% Indian/South Asian; 3.1% Chinese; 3.1% Vietnamese; 2.2%, other Asian 1.3%, American Indian.9%, Filipino .4%, Cambodian, and 0.4% Nepali. See Fig. 1 for a graphical representation.

5.2. Measures

All experimental conditions and measures were pilot tested for internal consistency and reviewed by a team of researchers familiar with the design and hypotheses.

5.2.1. Experimental conditions

Type of negativity distinguished between ambiguous negativity and clear hostility in live chat interactions. *Ambiguous negativity* is an umbrella term for negative behaviors that are open to interpretation (Cook et al., 2018; Karhulahti, 2016; Seering et al., 2017). Prior research suggests that interpreted in context they might not be negative (see Cruz et al., 2018). The *ambiguous negativity* condition included swear words without other-directed animosity. *Clear hostility* comprises behaviors that use exclusionary language (Cook et al., 2018, 2019) and other-directed aggressiveness. See supplemental materials.

Active vs. passive involvement was the extent to which community-specific rules imply that community members should be active in responding to negativity. Across conditions, rule sets had the same information regarding what constitutes negative behavior. The *active involvement* condition encouraged participants to act proactively and shape their community positively by calling out people who behave inappropriately. The *passive involvement* condition assigned participants a less active role by encouraging them to ignore or not engage with transgressors (e.g., haters or trolls). The *control* condition presented participants with rules which did not discuss the degree of involvement in responding to negativity. Rule sets were designed to resemble the format and phrasing of actual Twitch.tv channel rules to increase ecological validity. The language and phrasing of the rules were modeled after the researchers' experiences viewing live streams with rules, as well as prior examples such as Mihailova (2022). This was an intentional choice that prioritized rules design that would sound more natural over rules design that would fully minimize any non-manipulated differences among the conditions. While we strove to limit any differences in the rule sets outside those due to the manipulated variable of active vs. passive involvement (i.e., all other rules pertaining to what is allowed or not allowed are the same verbatim), we thought that encountering rules that resemble what actual Twitch.tv channel rules sound like would be better at eliciting hypothetical social identification from participants who use Twitch.tv already. See

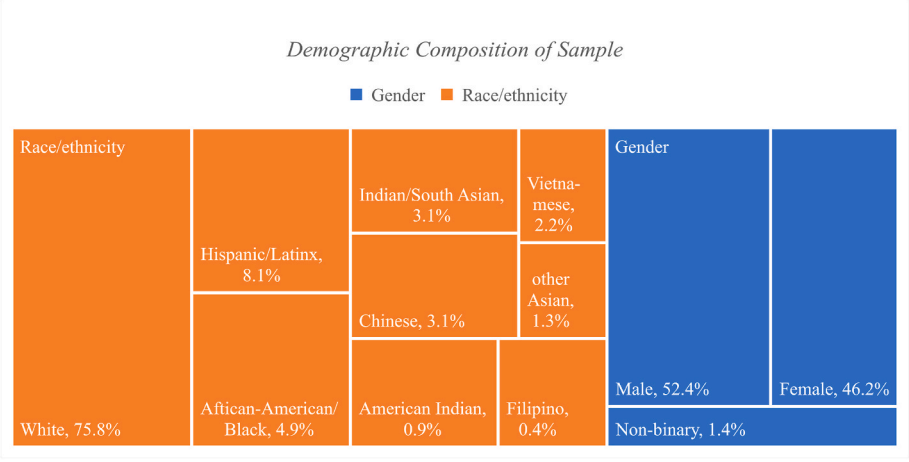

Fig. 1. Demographic Composition of Sample
Note: Chart areas are not proportional to prevalence of demographic characteristic; most prevalent categories scaled down for readability.

supplemental materials.

5.2.2. Manipulation checks

Perceived degree of involvement measured the effectiveness of the agency manipulation using one item on a 5-pt Likert type scale, "This community asks participants to hold each other accountable for their behavior in the chat." An ANOVA showed difference between perceived agency in the high agency condition ($M = 3.52$) and the low agency condition ($M = 3.14$), was significant, $p = .022$. The control condition ($M = 3.30$) and low agency conditions ($M = 3.14$) were not significantly different, $p = .249$.

Perceived appropriateness had five items (including one negatively worded item), such as "*This community would consider the chat messages highlighted in yellow to be hostile or offensive*" ($\alpha = 0.94$). An independent samples *t*-test showed that appropriateness was significantly lower in the clear negativity condition ($M = 3.11$) than the ambiguous negativity condition ($M = 4.44$), $t(210) = 12.30$, $p < .001$.

5.2.3. Moderator variable

Social identification. The measures of social identification were based on Chung's (2019) measures of social identification for online comments adapted from prior SIDE research (Henry, Arrow, & Carini, 1999). Participants reported how much they identified with live chat participants who were *not* breaking the rules. To assure participants understood which messages they were supposed to be evaluating, those were highlighted and color coded differently from the rest of the messages (see supplemental materials). Social identification had six items (including one negatively worded item) on a 5-pt Likert type scale, such as "*I can easily relate to these live chat participants*" ($\alpha = 0.86$).

5.2.4. Dependent variables

Likelihood of taking action was operationalized as four 5-pt Likert-type items measuring how likely participants would be to take each of the following actions with regard to the negative behavior they witnessed: "*Report the behavior privately to the streamer or a moderator*", "*Call out the behavior by confronting the person in the chat*", "*Ignore the behavior so as not to validate the person*" (reverse coded), "*Disregard the behavior because it is not problematic*" (reverse coded) ($\alpha = 0.75$).

Perceived group norms measured participants' perception of the local norms of the hypothetical community (e.g., what types of speech are allowed and appropriate; Chung, 2019). Participants answered four 5-pt Likert-type items, including one negatively worded item, about their impression of what was allowed based on the rules they saw, such as "*Racist, sexist, ableist, or otherwise offensive comments are not permitted in the live chat of this stream*" ($\alpha = 0.79$). Higher scores indicated a better perception that clearly negative behavior was not permissible in the channel.

5.2.5. Control variables

Prior experience with live streaming measured frequency of watching video game live streams on Twitch.tv and in general. Participants rated their frequency of use (Ajzen, 2002) of live streaming platforms (e.g., "*I watch live videos of others playing video games over live streaming platforms"*) and Twitch.tv in particular (e.g., "*I watch video game live streams on Twitch.tv"*) on a scale of 5 (*Multiple times a week*), 4 (*Weekly*), 3 (*Monthly*), 2 (*A few times a year*), 1 (*Never*) ($\alpha = 0.93$).

Chat activeness measured how much participants consider themselves as active in stream live chats in their prior live stream usage using 5-pt Likert-type five items, including two negatively worded items, such as "*I*

Table 1
Group Sizes and measured variable means by experimental group.

Active /Passive Involvement (Manipulated)	Type of Negativity (Manipulated)		Perceived Appropriateness of Messages (Measured)	Perceived degree of involvement (Measured)	Likelihood of taking action (Measured)	Perceived Group Norms (Measured)	Social Identification (Measured)
Control	Ambiguous	Mean	3.17	3.28	2.33	3.82	3.12
		N	33	33	35	34	33
		SD	.93	.77	.71	.89	.67
	Clear	Mean	4.54	3.30	2.99	3.89	3.41
		N	37	37	38	37	37
		SD	.58	.96	1.01	.89	.81
	Total	Mean	3.90	3.30	2.67	3.86	3.27
		N	70	70	73	71	70
		SD	1.02	.87	.94	.89	.75
Low	Ambiguous	Mean	3.18	3.19	2.16	3.92	3.05
		N	37	37	37	37	37
		SD	1.00	.78	.81	.86	.80
	Clear	Mean	4.47	3.23	2.76	3.91	3.36
		N	37	37	37	37	37
		SD	.61	.98	.86	.83	.89
	Total	Mean	3.83	3.21	2.46	3.92	3.20
		N	74	74	74	74	74
		SD	1.05	.88	.88	.84	.85
High	Ambiguous	Mean	2.98	3.36	1.94	3.86	3.16
		N	34	35	36	35	34
		SD	.89	.73	.62	.76	.80
	Clear	Mean	4.29	3.42	3.17	3.71	3.32
		N	34	34	34	34	34
		SD	.62	.71	.78	.97	.60
	Total	Mean	3.64	3.39	2.54	3.79	3.24
		N	68	69	70	69	68
		SD	1.01	.72	.93	.87	.70
Total	Ambiguous	Mean	3.11	3.28	2.14	3.87	3.11
		N	104	105	108	106	104
		SD	.94	.76	.73	.83	.75
	Clear	Mean	4.44	3.31	2.97	3.84	3.36
		N	108	108	109	108	108
		SD	.60	.89	.90	.89	.77
	Total	Mean	3.79	3.29	2.56	3.86	3.24
		N	212	213	217	214	212
		SD	1.03	.83	.92	.86	.77

tend to be active in the chat when I watch Twitch.tv live streams" ($\alpha = 0.90$).

See Table 1 for means and standard deviations, and Table 2 for correlation matrix.

6. Results

In this experiment, participants were exposed to one of six (two [clear vs. ambiguous negativity] by three [encouraging active involvement vs. encouraging passive involvement vs. control]) experimental conditions containing hypothetical channel rules and a researcher-generated live chat segment. Following the experimental manipulations, participants completed self-report survey measures of the study variables (i.e., likelihood of taking action, perceived group norms, social identification, as well as manipulation checks and participant characteristics). Table 3 summarizes our hypotheses and whether or not we found support for them. Below, we detail the statistical analyses of the survey responses conducted to test hypotheses.

A MANCOVA was run to predict likelihood of taking action based on type of negativity, active vs. passive involvement, and social identification. Prior experience and chat activeness were covariates. The MANCOVA indicated no overall differences among the active vs. passive involvement vs. control conditions (showing a lack of support for H2), Wilks $\Lambda = .99$, $F(4, 404) = 0.58$, $p = .678$, partial $\eta^2 = 0.006$, but there was an overall difference between the negativity conditions (in support of H1), Wilks $\Lambda = 0.76$, $F(4, 202) = 31.27$, $p < .001$, partial $\eta^2 = 0.236$. There was also an overall effect for social identification, Wilks $\Lambda = 0.95$, $F(2, 202) = 4.33$, $p = .014$, partial $\eta^2 = 0.041$.

Consistent with H3, interaction effects analyses detected a negativity by active vs. passive involvement interaction, Wilks $\Lambda = 0.96$, $F(4, 404) = 3.819$, $p = .024$, partial $\eta^2 = 0.036$. Between-subjects effects regarding likelihood of taking action showed the following effects: chat activeness, $F(1, 203) = 2.96$, $p = .087$, partial $\eta^2 = 0.014$; prior experience, $F(1, 203) = 5.28$, $p = .023$, partial $\eta^2 = 0.025$; social identification $F(1, 203) = 0.21$, $p = .644$, partial $\eta^2 = 0.001$; active vs. passive involvement, $F(1, 203) = 0.89$, $p = .412$, partial $\eta^2 = 0.009$; type of negativity, $F(1, 203) = 57.83$, $p < .001$, partial $\eta^2 = 0.222$. The active vs. passive involvement by type of negativity interaction effect was significant, $F(2, 203) = 2.16$, $p = .038$, partial $\eta^2 = 0.032$. Examining Fig. 2, the likelihood of taking action was higher across conditions when negativity was clear rather than ambiguous. Compared to the other conditions, participants in the active involvement condition were more likely to get involved when the negativity was clear and less likely to get involved when the negativity was ambiguous, supporting H3 (see Fig. 3).

H4 predicted an interaction effect between social identification and (a) type of negativity and (b) active vs. passive involvement in predicting likelihood of taking action. When predicting likelihood of taking action, the interaction between social identification and type of negativity was significant, $B = 0.63$, $SE = 0.24$, $p = .009$, providing support for H4a. The two-way interaction with social identification and active vs. passive involvement was not significant, $B = 0.14$, $SE = 0.14$, $p = .302$ (which showed a lack of support for H4b), nor was the three-way interaction, $B = -0.29$, $SE = 0.20$, $p = .136$, which did not support H5. Upon probing the significant interaction (Fig. 4), as degree of social identification increased the likelihood of taking action increased in cases of clear negativity. As degree of social identification increased, likelihood of taking action decreased for ambiguous negativity.

H6 and H7 predicted that group norms would be influenced by the two experimental conditions and associated with social identification. Between-subjects effects regarding perceived norms, showed the following effects: chat activeness, $F(1, 203) = 4.92$, $p = .028$, partial $\eta^2 = 0.024$; prior experience, $F(1, 203) = 6.11$, $p = .014$, partial $\eta^2 = 0.029$; social identification $F(1, 203) = 8.70$, $p = .004$, partial $\eta^2 = 0.041$; active vs. passive involvement, $F(1, 203) = 0.23$, $p = .791$, partial $\eta^2 = 0.002$; type of negativity, $F(1, 203) = 1.68$, $p = .195$, partial $\eta^2 = 0.008$. Neither type of negativity nor the type of involvement encouraged influenced perceived norms, so no support was found for H6 or H7. The active vs. passive involvement by type of negativity interaction effect was not significant, $F(1, 203) = 0.25$, $p = .777$, partial $\eta^2 = 0.002$. Overall, the results suggest that active vs. passive involvement and type of negativity, nor the interaction between the two, predicted perceived norms.

Table 3
Summary of hypotheses and findings.

#	Hypothesis (abbreviated phrasing)	Support found: Yes /No
H1	Clearer negativity will be positively associated with likelihood of taking action.	Yes
H2	More active (vs. passive) involvement will be positively associated with likelihood of taking action.	No
H3	Encouraging active (vs. passive) involvement will strengthen the relationship between type of negativity and likelihood of taking action.	Yes
H4a	Higher social identification will strengthen the relationship between type of negativity and likelihood of taking action.	Yes
H4b	Higher social identification will strengthen the relationship encouraging active (vs. passive) and likelihood of taking action.	No
H5	Likelihood of taking action will be highest when type of negativity is clear, rules encourage active (as opposed to passive) involvement, and social identification is high.	No
H6	Clearer negativity will be positively associated with correctly perceiving group norms.	No
H7	Encouraging active (vs. passive) involvement will be positively associated with correctly perceiving group norms.	No

7. Discussion

This experiment placed participants who were familiar with live streaming platforms within a hypothetical Twitch.tv channel chat with researcher-generated channel rules and live chat excerpts. Half the participants were shown clear negativity (a segment with exclusionary language) and the other half were shown ambiguous negativity (a segment with explicit language but no hostility). The degree to which the rules encouraged active vs. passive involvement was manipulated based on the rules they were exposed to, where the active involvement condition encouraged participants to be proactive when witnessing negativity, the passive involvement asked them to not engage with negativity, while the control condition had no rules pertaining to suggested involvement.

Results suggest being exposed to an excerpt with clear negativity increased participants' likelihood of calling that behavior out, while being exposed to rules that offered more active involvement did not. The more clearly the rules were violated, the more participants saw the need to react. As predicted, however, there was an interaction effect where participants in the active involvement condition were more likely to get involved when the negativity was clear and less likely to get involved when the negativity was ambiguous. Since the ambiguous negativity condition included swearing, which might be considered rude or inappropriate but was not hostile or rule violating, results suggest that rule

Table 2
correlations of non-manipulated variables.

Perceived degree of involvement	.192**					
Likelihood of taking action	.444**	.356**				
Perceived group norms	.071	.271**	.036			
Social identification	.101	.203**	.062	.230**		
Prior experience	−.081	−.032	−.132	.201**	.340**	
Chat activeness	−.089	.035	.031	−.069	.209**	.251**

**. Correlation is significant at the 0.01 level (2-tailed).

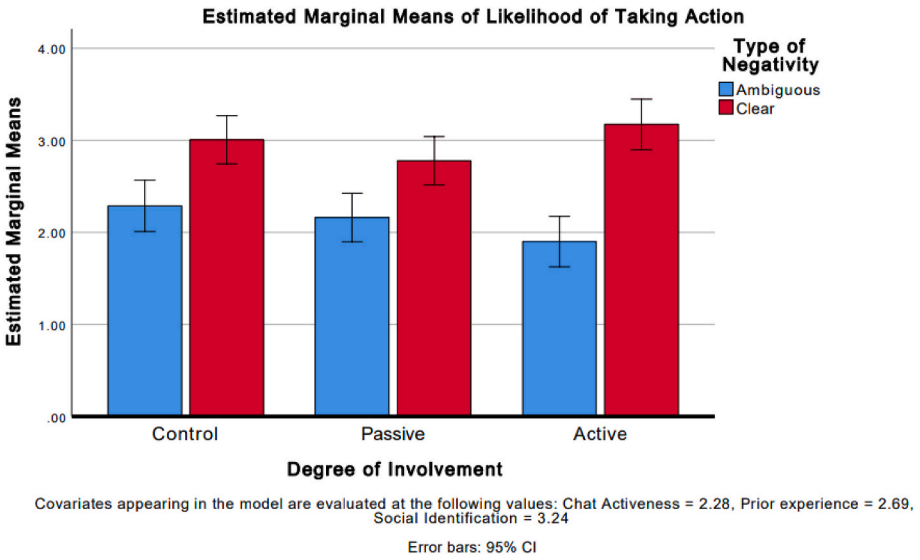

Fig. 2. *Likelihood of Taking Action by Experimental Condition* ($N = 211$).

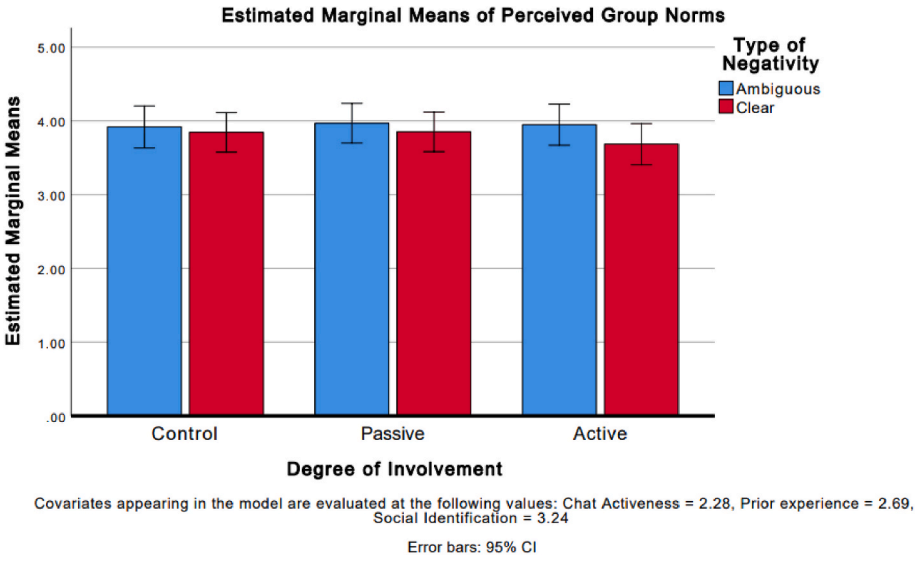

Fig. 3. *Perceived Group Norms by Experimental Condition* ($N = 211$).

sets that call on chat participants to get involved if they see something inappropriate make chat participants more capable of, and even more discerning about, when to apply the rules. This could be interpreted as evidence that when participants were encouraged to be involved actively, they showed a more nuanced understanding of the rules. This effect is consistent with the theoretical underpinnings of this study: the degree of active vs. passive involvement relates to the influence dimension of McMillan and Chavis's (1986) sense of community, which in turn is associated with conformity to group norms (Perfumi et al., 2019). A higher likelihood of getting involved when the rules encourage active involvement is consistent with SIDE's (Postmes et al., 1998) assertion that in anonymous conditions when the norms are made clear (e.g., by a streamer's rule set), individuals who identify with that community are motivated to conform to norms. Herein, reacting against rule violators. While the degree of negativity was an important factor in explaining when participants saw the behavior as violating rules, encouraging active involvement in chat participants clarified when intervention was needed. This evidence in favor of the SIDE model contributes to the theory by applying it to a novel medium (i.e., synchronous stream live chats) using hypothetical scenarios as manipulations. This serves both as testament to SIDE's applicability to such contexts and as proof of concept for an unusual methodological approach (i.e., using manipulated rules to attempt elicit identification and involvement with a hypothetical community).

Social identification showed a similar interaction effect. Type of negativity interacted significantly with social identification in predicting likelihood of taking action. Specifically, greater self-reported social identification increased likelihood of intervening when negativity was clear and decreased it when negativity was ambiguous. This is consistent with McMillan and Chavis's (1986) construct of sense of community because identification might give participants a greater sense of emotional connection, thus strengthening their sense of community and likelihood of enforcing group norms. While not directly examining the same variables, this finding is consistent with predictions outlined by Seering et al. (2018) that making salient the positive aspects of social identity might be able to reduce toxic behavior: those participants who did experience higher social identification were more discerning and reported higher likelihood of stepping in when rules were violated and lower likelihood of stepping in in ambiguous situations where rules were not actually violated, which suggests a willingness to combat negativity

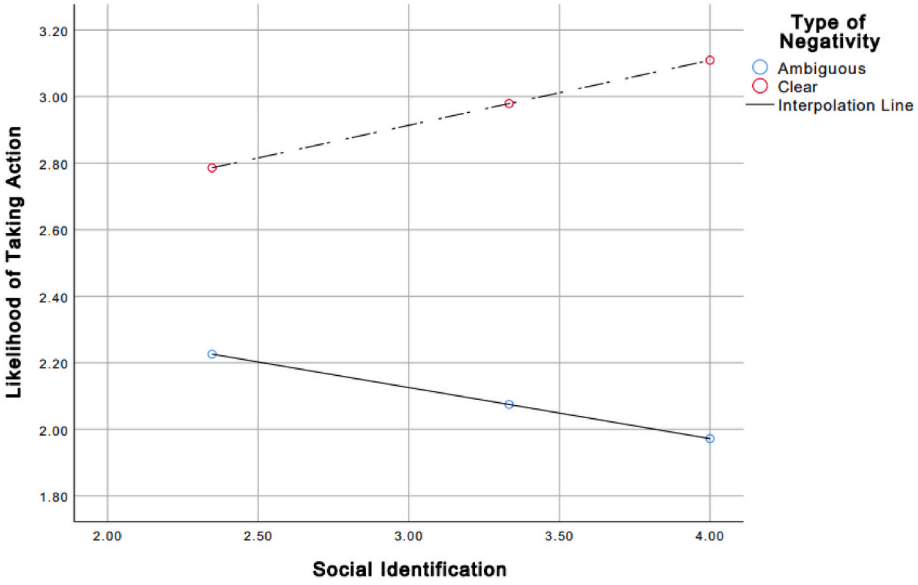

Fig. 4. Identification by type of negativity interaction on likelihood of taking action.

by stepping in while remaining sensitive to the local norms.

Against hypotheses, social identification did not interact with the degree of active vs. passive involvement nor was there a three-way interaction. Although providing an explanation for a lack of significant results is notoriously fraught, it is important to consider why degree of negativity but not the degree of active vs. passive involvement interacted with social identification. One explanation could be that because participants were asked about their identification with a hypothetical community (rather than a particular real online community), they might have defaulted to their general identification with live streaming viewers based on prior experience. Inducing different levels of identification based on small variations in hypothetical rules might not have been strong enough to produce effects. Additionally, perhaps Twitch.tv channel rules might not be integral in shaping social identification. Specific communities vary in how much active involvement they expect from members (Mihailova, 2022), but perhaps not directly because of the rules. Yet, this seems to contrast some past findings that found rules to be quite important in directing community action (e.g., Kowert, Botelho, & Newhouse, 2022; Myles et al., 2020). While rules might be an aspect of community identity in spaces like these, it is also possible that identity and community-specific experiences are formed more holistically by observing practices persistently (Chandrasekharan et al., 2018).

Counter to predictions, perceptions of group norms were not affected by the degree of active vs. passive involvement in rule sets. One potential reason for the lack of effect of active involvement on perceived norms is the norms in the experiment were too clear, so participants' perceptions lacked variability. Though the measures for perceived group norms performed consistently, those items focused predominantly on rules about exclusionary language and hostility, which is considered anti-normative in general or as platform-level rather than community-level norms (Chandrasekharan et al., 2018). Participants may have understood the norms irrespective of what rules and chats they were shown. Because an inclusion criterion was familiarity with the platform, the sample was limited to people who likely understood the norms. A finding that supports this explanation is the association between perceived group norms and prior experience with live stream viewership, used as a control variable. Although not specifically hypothesized, results suggest people with more experience were better at discerning what did and did not go against community rules. Outsiders to live-streaming might interpret chat behaviors differently (Mihailova, 2022). Future work should explore how group norm perceptions are formed and whether they vary between communities, games, and platforms.

Overall, findings indicate that more clear-cut norm violations are more easily identifiable and more readily responded to. Active involvement and higher social identification each made people more willing to react to clear-cut norm violations and less likely to react to ambiguous negativity that was not hostile. By contrast, perceptions of group norms appeared to be stable across conditions, affected only by variables related to participants' prior experiences with live streams, but not by the rules or chat excerpts in the experiment.

7.1. Limitations

This study used new measures and manipulations, with a limited convenience sample, a cross-sectional design, and pertaining to a hypothetical community. Only people with experience viewing live streams were recruited, thus findings likely do not generalize to either the population or live stream viewers as a whole. This study showed small effect sizes overall and cannot be considered as a direct solution to negativity in chat.

Another considerable limitation of this study is its hypothetical nature – both in terms of the community participants are encouraged to identify with and in terms of participants' responses (i.e., participants self-report their likelihood of stepping in for the hypothetical situation presented to them, which is not the same as putting them in such a situation and observing whether they would actually step in). We sought to mitigate these limitations through our research design: we followed previous work (Chung, 2019; Rösner & Krämer, 2016) when attempting to create identification in a zero-history hypothetical community and we strove to design realistic rules and chat messages to improve ecological validity and make participants' self-reported likelihood to respond a more realistic reflection of what they might actually do in such a situation. However, these limitations cannot be fully eliminated because of the nature of the design.

7.2. Conclusions

All the newly developed measures showed acceptable to good reliability and could be used in future work in ecologically valid settings. Participants could think of a particular Twitch.tv channel community, and answer questions pertaining to the specific norms of their chosen community and the degree to which it expects active or passive involvement. Studying real communities might better explain what role

(if any) rules might play in shaping identity and involvement, especially over time with a longitudinal design. When it comes to negativity, it might be helpful to design manipulations with varying degrees of ambiguity to capture the range of ambiguous behaviors encountered in live chats (Mihailova, 2022). Sampling beyond live stream viewers could capture variation in how outsiders differ from experienced live stream viewers.

Understanding what shapes community members' willingness to step in and how that might be shaped by rules has practical implications for community management. Twitch.tv channel rules could be studied similarly to how Fiesler, McCann, Frye, and Brubaker (2018) explored community-generated rules on Reddit. If rules are indeed an important component of fostering a sense of responsibility and active involvement, this would be important information to moderators and community managers. This relationship could potentially have important practical implications if explored further – the present study showed that when active involvement is encouraged and when social identification is activated, participants would be more willing to respond appropriately to negativity; future work should establish more directly whether rules and other methods might be effective ways of activating positive social identity to accomplish that in practice. That line of research has the potential to benefit online community managers and members by discovering and promoting evidence-based practices to leverage rules design and ordinary user involvement against negativity. Overall, this research contributes to understanding alternative ways to foster a sense that one's active involvement is expected and valued and empower members to respond to hostile behavior in contextually appropriate ways without simultaneously increasing moderator burden.

Author note

Work performed at Department of Communication Studies, University of Kansas.

The article processing charges related to the publication of this article were supported by The University of Kansas (KU) One University Open Access Author Fund sponsored jointly by the KU Provost, KU Vice Chancellor for Research, and KUMC Vice Chancellor for Research and managed jointly by the Libraries at the Medical Center and KU - Lawrence.

We have no known conflicts of interest to disclose.

All data will be made available upon reasonable request by the first author.

Author contributions

TM: Conceptualization; Data analysis; Data curation; Methodology; Project administration; Resources; Writing - original draft; Writing - review & editing.

JH: Supervision; Data Validation; Data Visualization; Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2023.100358.

References

Ajzen, I. (2002). *Constructing a TPB questionnaire: Conceptual and methodological considerations*.

Bergstrom, K. (2011). "Don't feed the troll": Shutting down debate about community expectations on Reddit. com. *First Monday, 16*(8). https://doi.org/10.5210/fm.v16i8.3498

Blight, M. G. (2016). *Relationships to video game streamers: Examining gratifications, parasocial relationships, fandom, and community affiliation online. PhD thesis*. US: University of Wisconsin-Milwaukee. https://dc.uwm.edu/etd/1255.

Cai, J., & Wohn, D. Y. (2019). What are effective strategies of handling harassment on Twitch? Users' perspectives. In *Conference companion publication of the 2019 on computer supported cooperative work and social computing* (pp. 166–170). https://doi.org/10.1145/3311957.3359478

Chandrasekharan, E., Samory, M., Jhaver, S., Charvat, H., Bruckman, A., Lampe, C., … Gilbert, E. (2018). The internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–25. https://doi.org/10.1145/3274301

Chen, C., & Lin, Y. (2018). What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement. *Telematics and Informatics, 35*(1), 293–303. https://doi.org/10.1016/j.tele.2017.12.003

Chung, J. E. (2019). Peer influence of online comments in newspapers: Applying social norms and the social identification model of deindividuation effects (SIDE). *Social Science Computer Review, 37*(4), 551–567. https://doi.org/10.1177/0894439318779000

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*(6), 1015. https://doi.org/10.1037/0022-3514.58.6.1015

Cook, C., Conijn, R., Schaafsma, J., & Antheunis, M. (2019). For whom the gamer trolls: A study of trolling interactions in the online gaming context. *Journal of Computer-Mediated Communication, 24*(6), 293–318. https://doi.org/10.1093/jcmc/zmz014

Cook, C., Schaafsma, J., & Antheunis, M. (2018). Under the bridge: An in-depth examination of online trolling in the gaming context. *New Media & Society, 20*(9), 3323–3340. https://doi.org/10.1177/14614448177485

Cruz, A. G. B., Seo, Y., & Rex, M. (2018). Trolling in online communities: A practice-based theoretical perspective. *The Information Society, 34*(1), 15–26. https://doi.org/10.1080/01972243.2017.1391909

Diwanji, V., Reed, A., Ferchaud, A., Seibert, J., Weinbrecht, V., & Sellers, N. (2020). Don't just watch, join in: Exploring information behavior and copresence on Twitch. *Computers in Human Behavior, 105*. https://doi.org/10.1016/j.chb.2019.106221

Fiesler, C., McCann, J., Frye, K., & Brubaker, J. R. (2018). Reddit rules! characterizing an ecosystem of governance. *Twelfth International AAAI Conference on Web and Social Media, 12*(1). Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/15033.

Gillepsie, T. (2017). Regulation of and by platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media*. SAGE Publications Ltd. https://doi.org/10.4135/9781473984066.

Guarriello, N.-B. (2019). Never give up, never surrender: Game live streaming, neoliberal work, and personalized media economies. *New Media & Society, 21*(8), 1750–1769. https://doi.org/10.1177/1461444819831653

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research, 6*(2), 215–242. https://doi.org/10.1515/jplr.2010.011

Henry, K. B., Arrow, H., & Carini, B. (1999). A tripartite model of group identification: Theory and measurement. *Small Group Research, 30*(5), 558–581. https://doi.org/10.1177/104649649903000504

Hilvert-Bruce, Z., Neill, J. T., Sjöblom, M., & Hamari, J. (2018). Social motivations of live-streaming viewer engagement on Twitch. *Computers in Human Behavior, 84*, 58–67. https://doi.org/10.1016/j.chb.2018.02.013

Hu, M., Zhang, M., & Wang, Y. (2017). Why do audiences choose to keep watching on live video streaming platforms? An explanation of dual identification framework. *Computers in Human Behavior, 75*, 594–606. https://doi.org/10.1016/j.chb.2017.06.006

Jhaver, S., Birman, I., Gilbert, E., & Bruckman, A. (2019). Human-machine collaboration for content regulation: The case of Reddit Automoderator. *ACM Transactions on Computer-Human Interaction, 26*(5), 1–35. https://doi.org/10.1145/3338243

Karhulahti, V.-M. (2016). Prank, troll, Gross and Gore: Performance issues in esports live-streaming. *Proceedings of 1st international joint conference DiGRA and FDG*.

Kowert, R., Botelho, A., & Newhouse, A. (2022). *Breaking the building blocks of hate: A case study of minecraft servers*. ADL Center of Technology and Society. https://www.adl.org/resources/report/breaking-building-blocks-hate-case-study-minecraft-servers.

London, T. M., Crundwell, J., Eastley, M. B., Santiago, N., & Jenkins, J. (2019). Finding effective moderation practices on Twitch. In *Digital ethics* (pp. 51–68). Routledge.

Marvin, L. E. (1995). Spoof, spam, lurk, and lag: The aesthetics of text-based virtual realities. *Journal of Computer-Mediated Communication, 1*(2), 122. https://doi.org/10.1111/j.1083-6101.1995.tb00324.x

Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media+ Society, 5*(3). https://doi.org/10.1177/2056305119862641

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology, 14*(1), 6–23. https://doi.org/10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I

Mihailova, T. (2022). Navigating ambiguous negativity: A case study of Twitch. Tv live chats. *New Media & Society, 24*(8), 1830–1851. https://doi.org/10.1177/1461444820978999

Myles, D., Benoit-Barné, C., & Millerand, F. (2020). 'Not your personal army!' Investigating the organizing property of retributive vigilantism in a Reddit collective of websleuths. *Information, Communication & Society, 23*(3), 317–336. https://doi.org/10.1080/1369118X.2018.1502336

Narassiguin, A., & Garnès, V.. The influence of COVID-19 on Twitch audience: How lockdown measures affect live streaming usage?. Retrieved from https://upfluence-common.s3.amazonaws.com/Covid19_Twitch.pdf.

Perfumi, S. C., Bagnoli, F., Caudek, C., & Guazzini, A. (2019). Deindividuation effects on normative and informational social influence within computer-mediated-communication. *Computers in Human Behavior, 92*, 230–237. https://doi.org/10.1016/j.chb.2018.11.017

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries?: SIDE-effects of computer-mediated communication. *Communication Research, 25*(6), 689–715. https://doi.org/10.1177/009365098025006006

Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin, 27*(10), 1243–1254. https://doi.org/10.1177/01461672012710001

Rebollo-Catalan, A., & Mayor-Buzon, V. (2020). Adolescent bystanders witnessing cyber violence against women and girls: What they observe and how they respond. *Violence Against Women, 26*(15–16), 2024–2040. https://doi.org/10.1177/1077801219888025

Rosenthal, H. M., & Belmas, G. I. (2021). Cyber-recapitulation? What online games can teach social media about content management. *Jurimetrics, 61*(3), 331–378. lib.ku.edu/login?url=https://www.proquest.com/scholarly-journals/cyber-recapitulation-what-online-games-can-teach/docview/2568315100/se-2.

Rösner, L., & Krämer, N. C. (2016). Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. *Social Media + Society, 2*(3), 1–13. https://doi.org/10.1177/2056305116664220

Seering, J., Kraut, R., & Dabbish, L. (2017). Shaping pro and anti-social behavior on Twitch through moderation and example-setting. *CSCW '17 Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. https://doi.org/10.1145/2998181.2998277

Seering, J., Ng, F., Yao, Z., & Kaufman, G. (2018). Applications of social identity theory to research and design in computer-supported cooperative work. *Proceedings of the ACM on human-computer interaction, 2*(CSCW), 1–34. https://doi.org/10.1145/3274771

Seering, J., Wang, T., Yoon, J., & Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New Media & Society, 21*(7), 1417–1443. https://doi.org/10.1177/1461444818821316

Sjöblom, M., & Hamari, J. (2016). Why do people watch others play video games? An empirical study on the motivations of Twitch users. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2016.10.019

Suh, A., & Wagner, C. (2013). Factors affecting individual flaming in virtual communities. In *2013 46th Hawaii international conference on system sciences*. https://doi.org/10.1109/HICSS.2013.230

Taylor, T. L. (2018). *Watch me play: Twitch and the rise of game live streaming*. Princeton University Press.

Teneketzi, K. (2021). Impoliteness across social media platforms: A comparative study of conflict on YouTube and Reddit. *Journal of Language Aggression and Conflict*. https://doi.org/10.1075/jlac.00066.ten

Wohn, D. Y., & Freeman, G. (2020). Audience management practices of live streamers on Twitch. *ACM International conference on interactive media experiences*, 106–116. https://doi.org/10.1145/3391614.3393653

Yu, E., Jung, C., Kim, H., & Jung, J. (2018). Impact of viewer engagement on gift-giving in live video streaming. *Telematics and Informatics, 35*(5), 1450–1460. https://doi.org/10.1016/j.tele.2018.03.014