Media and Communication
2025 • Volume 13 • Article 8676
https://doi.org/10.17645/mac.8676

ARTICLE

Open Access Journal

Sympathy for the Devil: Serial Mediation Models for Toxicity, Community, and Retention

Mingxuan Liu [1], Jack Lipei Tang [2], and Dmitri Williams [3]

[1] Department of Communication, University of Macau, China
[2] Department of Advertising and Public Relations, The University of Alabama, USA
[3] Annenberg School for Communication and Journalism, University of Southern California, USA

Correspondence: Mingxuan Liu (mxliu@um.edu.mo)

Submitted: 14 May 2024 **Accepted:** 24 July 2024 **Published:** 13 February 2025

Issue: This article is part of the issue "Digital Games at the Forefront of Change: On the Meaningfulness of Games and Game Studies" edited by Felix Reer (University of Muenster), Teresa de la Hera (Erasmus University Rotterdam), and Salvador Gómez-García (Complutense University of Madrid), fully open access at https://doi.org/10.17645/mac.i460

Abstract

Disruptive behaviors in online gaming communities are a growing concern, affecting player experience, retention, and well-being. While previous research has primarily focused on the victims' experiences, this study examines the psychological mechanisms underlying the attitudinal and behavioral responses to both encountering disruptive behaviors and being flagged for such behaviors, as well as the effects on retention. The study retrieved longitudinal telemetry records of player reporting and gameplay data from the North American server of a popular competitive player vs. player multiplayer online game, coupled with a psychometric survey of a randomly selected sample of 1,217 players. Based on the rejection-disidentification model, this research identifies a shared pathway for both reporting and being reported for disruptive behavior. Our findings support a serial mediation model where both experiences are linked to decreased player engagement. This reduced engagement, reflected in diminished participation in game battles over time, is mediated by perceived discrimination and a reduced sense of community. Moreover, drawing on the concept of procedural justice from the group engagement model, the study delineates unique pathways for the disengagement process for reporters and those reported. Being flagged for disruptive behavior leads to a significant drop in sustained engagement through a decreased sense of community, which is not the case for reporting disruptive behavior. The article concludes with a discussion of the theoretical and practical implications of these findings.

Keywords

disruptive gaming behavior; gaming toxicity; online gaming community; online video games; player engagement; procedural justice; sense of community

1. Introduction

There is no winner in gaming toxicity. Toxic behavior occurs when players break co-existence rules and act in antisocial ways (Neto et al., 2017). This behavior may not always be intentional (Kordyaka et al., 2020). Players might engage in toxic behavior because they perceive it as the norm of the game based on their own experiences, or they might use toxic behavior as a coping mechanism for stress induced by gameplay (Neto et al., 2017). Regardless of their intentions, gaming toxicity is disruptive. It undermines a positive gaming experience not only for those directly encountering it but also for everyone involved—victims, bystanders, the gaming industry, and, rarely considered, the perpetrators themselves (Anti-Defamation League, 2023). Recent statistics indicate that approximately 75% of online multiplayer gamers in the US have encountered hate and harassment in the past six months, among whom 9% reported depressive or suicidal thoughts (Anti-Defamation League, 2023).

Research on gaming toxicity has identified a recurring theme: social identity. Networked games, such as online multiplayer games, serve as a nexus where technology, identity, society, and various forms of inequality, power, and discrimination intersect (Gray, 2012). Individuals develop their social identity and sense of identification and belonging to the mainstream group from the identity information they receive from their interactions within the group. This social identity approach, comprising social identity theory (SIT; Tajfel et al., 1979) and self-categorization theory (SCT; Turner et al., 1987), underpins the rejection-disidentification model (RDIM; Jasinskaja-Lahti et al., 2009) and the group engagement model (GEM; Tyler & Blader, 2003).

RDIM was developed to understand the racial and ethnic discrimination experienced by immigrants during their acculturation process, and studies have confirmed a mediation mechanism wherein decreased national identification mediates the impact of discrimination on various outcomes (e.g., Jasinskaja-Lahti et al., 2009, 2018). This study extends RDIM to virtual environments, exploring the acculturation experiences of video game players, particularly the influence of discrimination on their sense of community and ongoing engagement. Online games, albeit digital, parallel the acculturation experiences described in RDIM, as they mirror its foundational mechanism. Initially engaging with a video game is akin to immigrants entering a new country, where veteran players often represent the local majority, and each game, like a country, possesses distinct norms and cultural frameworks. Similar to the acculturative stress faced by immigrants, players encounter a cultural adaptation process and may experience stress upon integrating into an online multiplayer gaming environment. Immigrants are subjected to discrimination based on attributes such as race and ethnicity. Similarly, in the digital realm, where diverse social roles and identities converge (Gray, 2012), players may face discrimination based on racial and gender cues, as well as skill cues and gameplay styles (Nguyen et al., 2022). Identifying this similarity of individuals' acculturation experiences, this study tests whether the mediation mechanism identified in RDIM applies in the digital world, wherein a decreased sense of community mediates the impact of discrimination on player engagement.

GEM posits that the experience of unfair treatment within groups imparts critical identity information, potentially influencing individuals' motivation to maintain a positive group identity. This, in turn, mediates the relationship between perceived fairness and group engagement. RDIM enhances GEM's framework by interpreting perceived discrimination as a manifestation of unfair treatment. This study leverages RDIM and GEM to identify two serial mediation mechanisms in understanding whether and how involvement in gaming toxicity leads to perceived discrimination, which then diminishes players' sense of belonging to the

gaming community and their sustained engagement with the game. Moreover, drawing insights from GEM, this study posits that the inclusion of an in-game reporting tool can transform the power dynamics between perpetrators and victims (Reid et al., 2022). Leveraging GEM allows this study to outline potentially unique psychological paths for both victims and perpetrators (see Figure 1 for the conceptual framework).

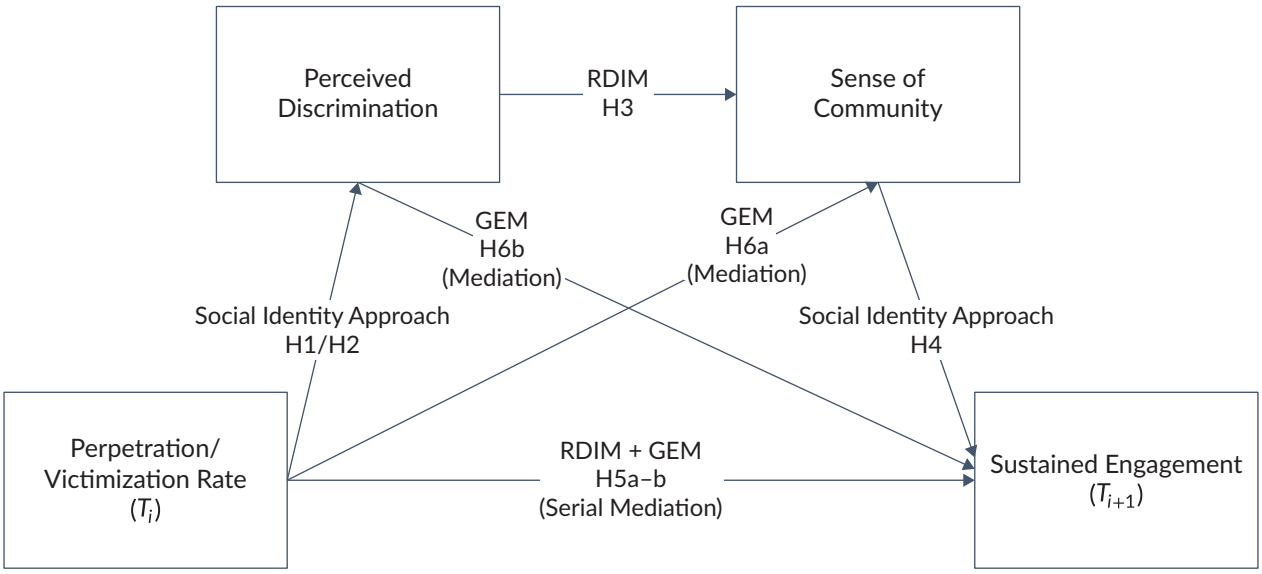


Figure 1. Conceptual framework.

2. Disruptive Gaming Behavior and Perceived Discrimination

Discrimination, as defined by Passmore and Mandryk (2020), represents a threat to an individual's inherent legitimacy and agency across various social identity categories. Understanding the emergence of discrimination is critical for effectively addressing it. The social identity approach has been extensively applied in understanding discrimination within and between groups, both online and offline (e.g., Adachi et al., 2015; Trepte & Loy, 2017). The two theories under the social identity approach, SIT and SCT, are intrinsically connected in their foundational premises and processes related to understanding intergroup relations and group dynamics. This study primarily examines the group processes associated with discrimination and its effects. According to SIT, social identity constitutes part of the self-concept derived from membership in social categories. It can be positive or negative, and individuals typically strive to enhance their social identity by affiliating with certain groups and valuing these associations (Trepte & Loy, 2017). Beyond self-categorization, individuals undergo ongoing social comparison with a reference group, often resulting in perceptions of the outgroup as inferior. This can fuel intergroup competition and conflict, even in the absence of direct rivalry, but for symbolic resources such as ranking, prestige, and positive group identity. For instance, in online games where players from different social groups form teams and compete, they strive to positively differentiate their social group from others by winning or strategically leveraging favorable statistics to maintain a positive social identity (Adachi et al., 2015).

SCT shares many ideas with SIT but also introduces uniquely formulated processes and predictions that offer distinct insights into discrimination. For example, SCT posits that the rich situational cues in media can make certain social categories more salient than others, thus influencing media users' perceptions and behaviors

(Trepte & Loy, 2017). When certain social identities become salient, especially in computer-mediated settings devoid of verbal cues, individuals' self-perception can be dominated by their social identity, leading to a phenomenon termed depersonalization. Depending on the level of depersonalization, individuals may conform to group norms, good or bad. Previous research has argued that game affordances, such as diverse in-game communication channels, can make players' social groups salient to others (Nguyen et al., 2022). For instance, in games like Call of Duty, players can display their national flags beside their names, making their national identity salient to others. Voice channels also reveal gender cues, potentially leading to gender harassment in online gaming. In addition to gender, race, and ethnicity-based discrimination, people can also possess disadvantages based on skills and age. In the realm of online video games, while some players view all others as part of their ingroup, others differentiate based on factors such as team affiliations and demographic indicators. This blending of online and offline identities can catalyze disruptive behaviors rooted in underlying prejudices and discrimination. Consequently, discrimination in online games is conveyed through various disruptive behaviors and perceived by those impacted. As such, it is hypothesized that:

H1: Incidents of reporting disruptive gaming behavior positively relate to perceived discrimination.

While reports of disruptive behavior are often submitted due to experienced discrimination, the impact extends beyond the feelings of the victims. Paradies (2006) highlights the importance of distinguishing between objective encounters of discrimination and its subjective interpretation—or perceived discrimination. Such perceived discrimination, which can also be experienced by those being reported, suggests rejection or exclusion from social groups that individuals aspire to be a part of, thus undermining group members' fundamental needs for social inclusion and acceptance.

In-game tools, like a complaint system, empower players facing toxicity to take immediate action. Recent research indicates that players who utilize these tools experience enhanced feelings of control, along with increased social and emotional support (Reid et al., 2022). However, reporting other players is a low-cost exercise and so is not infallible and can be a form of toxicity in itself. Those who are reported—whether justified or not—may experience a loss of control upon receiving an in-game warning. This notification can serve as a marker of potential or real social exclusion from the gaming community and signals that their behavior is viewed as disruptive by community norms. Given these dynamics, it is hypothesized that:

H2: Incidents of being reported for disruptive behavior positively relate to perceived discrimination.

3. Perceived Discrimination and Sense of Community

Perceived discrimination can obstruct the development of a sense of belonging to a superordinate in-group among certain members, as outlined by RDIM. The sense of belonging and identification with a group is fundamentally influenced by the identity cues provided by the group itself (Tyler & Blader, 2003). RDIM-based studies have shown that experiences of injustice and discrimination lead to a decline in identification with a larger group identity. This process, known as disidentification, occurs when individuals withdraw their loyalty and commitment from a group they previously identified with, driven by perceived barriers to developing a sense of community belonging (Jasinskaja-Lahti et al., 2018). In the context of the gaming community, the way individuals evaluate their experiences within the group can either motivate or deter them from maintaining a favorable group identity. Discriminatory experiences specifically convey to

individuals that they are not valued within the group, thereby discouraging identification with the broader gaming community. These processes suggest that perceived discrimination conveys negative identity information, which can lead to disidentification. Thus, it is predicted that:

H3: Perceived discrimination will be negatively associated with players' sense of community.

4. Sense of Community and Sustained Engagement

Sense of community is the perception of being part of a mutually supportive network of relationships (Sarason, 1974). Group membership is a fundamental aspect of social life, as it fulfills the basic human need to belong (Lind & Tyler, 1988). A supportive community not only meets this basic need but also fosters an ecosystem that enables members to lead fulfilling lives. Such a community provides a supportive network that members can rely on for psychological significance and identification, thus fostering a sense of community. Low identification with the community may lead to disengagement behaviors, such as reduced time investment, socialization, or cooperation (Zagenczyk et al., 2013).

This process is also evident in online gaming environments. Socialization is consistently identified as a major motivation for individuals to engage in online games (Ryan et al., 2006). Players enjoy participating in social activities within these computer-mediated environments to experience a sense of community and build social capital (Tseng et al., 2015). Indeed, previous research indicates that a player's sense of community mediates the relationship between their in-game social network and their intention to continue playing (Tseng et al., 2015). This is because the development of a sense of community often results from repeated social interactions. Functionally, in multiplayer games, players frequently form teams to complete missions or achieve common goals. Affectively, players unite over shared interests and often depend on each other for feedback and support, illustrating the mutual interdependence among players (Rovai, 2002). This interdependence can deter players from leaving the game, as doing so would mean losing contact with community members to whom they feel connected (Tseng & Teng, 2014). Therefore, it is hypothesized that:

H4: Players' sense of community will be positively associated with their sustained engagement.

5. Common and Distinct Routes to Disengagement for Reporters and Reported

In the context of online gaming, both reporting and being reported for disruptive behavior may lead to perceived discrimination, subsequently decreasing players' sense of community and dampening engagement. Although previous research supports each step of this mediated path, a comprehensive evaluation of the entire pathway has been limited, likely due to challenges in accessing matched in-game report data and players' psychometric self-reports. Utilizing data from the competitive multiplayer online video game World of Tanks (WoT), this study tests a serial mediation model:

H5a: Increased incidents of being reported will be negatively associated with sustained engagement, serially mediated by increased perceived discrimination and decreased sense of community.

H5b: Increased incidents of reporting will be negatively associated with sustained engagement, serially mediated by increased perceived discrimination and decreased sense of community.

In addition to the common routes through which both perpetrators and victims of disruptive behavior may perceive discrimination, different pathways to disengagement arise due to their unique experiences. GEM suggests that procedural justice in social settings plays a crucial role in motivating individuals to engage with groups. Perceived fairness of group procedures transmits vital identity-related information, providing a sense of identity security. When group decisions are perceived as being made through fair procedures, individuals are more likely to feel that their identities are securely associated with the group, encouraging comfortable psychological and behavioral engagement. In addition, research on procedural justice has demonstrated that individuals greatly value having a voice in the decision-making process. This opportunity has both interpersonal and value-expressive significance. The presence of a voice tends to make a procedure seem fairer, even if it does not directly influence the decision outcome. This perception stems from the belief that having a voice indicates that the group's authority respects individual members, acknowledging their values and arguments. During decision-making, group members focus on whether their concerns and needs are respected, independently of the actual influence their voice has on the final decision. The ability to exercise one's voice acts as a visible marker of group membership, reinforcing one's place within the group. Therefore, procedural justice is a foundational antecedent of group members' identities and their engagement with the group.

In addition to GEM, the exit, voice, and loyalty model (Hirschman, 1970) further elucidates the dynamics of group participation. This model suggests that users faced with dissatisfaction typically react in one of two ways: they may choose to exit or use their voice. While exit represents a withdrawal from the current situation, voice signifies an attempt to change it. If the product or service does not improve despite voiced concerns, users may exit and choose an alternative. Conversely, if their voices are acknowledged and lead to changes, users are likely to remain loyal.

This framework is particularly relevant in the setting of online video games, where both reporters and reported players encounter unique challenges related to discrimination and group dynamics. RDIM suggests that experiencing discrimination can be devastating to players' sense of social self, potentially leading to psychological and behavioral distancing from group membership. However, for reporters, there exists an alternative pathway. Although reporters are likely motivated by experiences of discrimination, their access to in-game reporting systems ensures that their grievances are heard and addressed, preventing damage to their identification with the gaming community. Conversely, being reported for disruptive behavior may directly impact the reported players' sense of community, leading to decreased engagement. This loss is tied to the punitive repercussions and social isolation that reported players often face, eroding their connection to the community. Based on these dynamics and drawing on the principles of procedural justice and the exit, voice, and loyalty model, we predict:

> H6a: Being reported for disruptive behavior is negatively associated with sustained engagement, mediated by a decreased sense of community.

> H6b: Being reported for disruptive behavior is negatively associated with sustained engagement, mediated by a decreased sense of community, which is not the case for reporting disruptive behavior.

6. Method

6.1. Data

The study site was WoT, a popular team-based vehicle simulator that has attracted over 160 million users worldwide. WoT features tank combat between two teams of up to 15 players. It differs from most other team-based shooters in that the teams are not matched by player skill ratings, but by tank types and tiers. Thus, players in WoT have a higher than usual chance to be in teams with disparate skill levels, which can cause lopsided results and potentially more emotional swings. WoT players can make complaints up to 10 times per day. Being reported five times can lead to permanent restriction from the game. The reporting system comprises four categories: inappropriate behavior in chat, unsportsmanlike conduct, offensive nicknames or clan names, and inaction/bot behavior. A player may be reported for multiple offenses simultaneously, with these categories not being mutually exclusive. Observations suggest that reports made during battles are often hastened, with players possibly selecting a category arbitrarily due to time constraints. This potential for overlap and inconsistency in category selection has led us to analyze the totality of complaint incidents to collectively assess different types of toxic behavior.

In collaboration with Wargaming, the operator of the game, we unobtrusively collected player-level report data and gameplay from the North American server, spanning February 2019 to June 2019. The data were anonymized before reaching the research team. To further understand the psychometric profiles of players, an online survey was randomly distributed to active WoT players in April 2019, achieving a response rate of 20.6% with 2,011 participants. We then matched the behavioral data with the survey responses using a unique, one-way hashed key. After excluding invalid responses, the dataset comprised 1,217 unique participants with 6,085 repeated observations.

6.2. Measures

The Perpetration rate was computed by dividing the number of reports a player received in a given month by the total number of battles that player participated in during the same month ($M = 0.011$, $SD = 0.025$). This calculation allowed us to distinguish between players who were frequently reported due to high levels of activity from those who were seldom active but disproportionately reported. The Victimization rate was calculated by dividing the number of times a player reports others for disruptive behavior in a given month by the total number of battles they participated in during that month ($M = 0.026$, $SD = 0.062$).

Perceived discrimination was assessed using the nine-item everyday discrimination scale (Williams et al., 1997), adapted for the WoT context. Players were prompted with the question, "In your day-to-day WoT gameplay, how often do any of the following things happen to you?" Responses were scored on a scale from one to six, where higher scores indicate a greater perception of discrimination. The items included experiences such as being treated with less courtesy than others, being insulted or called names, and being threatened or harassed. The responses were averaged to create an index representing the level of perceived discrimination during gameplay (*Cronbach's alpha* = 0.88, $M = 2.10$, $SD = 0.84$).

The sense of community among players was evaluated using a four-item scale developed by Kim (2011) and adapted for the WoT context. Sample questions included: "Even though we were physically in different

locations, I still felt I was part of a group of friends in the game." Responses were given on a scale from one to five, with higher scores indicating a stronger sense of community. The average of these responses was used to create an index reflecting the community feeling among players (*Cronbach's alpha* = 0.79, *M* = 3.10, *SD* = 0.96). Sustained engagement measured the number of battles a player engaged in each month following an occurrence of perpetration or victimization (*M* = 262.87, *SD* = 277.49).

We incorporated several covariates into our model to account for potential confounding variables. These included age (*M* = 43.17, *SD* = 16.29), education level (median: associate degree), and income level (median range: $50,000 to $74,999). Given the unbalanced demographics in terms of gender (N_{male} = 1,184, N_{other} = 33) and race (N_{white} = 1,007, N_{other} = 210), both variables were coded categorically. The majority group was assigned a value of one, and all other groups a value of zero. Additionally, in-game rating (*M* = 4,302, *SD* = 1,718.32) was used as a covariate, which is an indicator of performance within WoT. This rating, retrieved from the WoT North American server, is a composite score developed by a specific WoT algorithm that considers the number of battles played, win rate, and average tank level.

To further refine our analysis, players' motivations, which might influence their in-game behavior and sense of community, were controlled. These motivations were assessed using the trojan player typology (Kahn et al., 2015), a validated five-point Likert scale adapted to reflect the current game context. The socialization motivation was measured by items such as "it's important to me to play with a tightly knit group" (α = 0.81, *M* = 3.20, *SD* = 1.20). The competence motivation was measured by items such as "winning is a big reason for me to play World of Tanks" (α = 0.77, *M* = 3.60, *SD* = 0.92).

In modeling sustained engagement, we included the number of battles each player participated in during the previous month as a lagged variable. This approach not only captures the temporal dynamics of players' activity levels but also helps adjust for autocorrelation, thereby providing more accurate and reliable estimates of players' activity trends and enhancing the predictive validity of the findings.

6.3. Analyses

This study employed a longitudinal design, using lagged independent variables from February to May 2019 to predict sustained engagement in the subsequent months from March to June 2019. Mediators were assessed through a survey conducted in April 2019. Due to the skewness of the variables and for ease of interpretation, all variables except for categorical ones were natural log-transformed. Consequently, a one percent change in the predictors translates to a one percent change in the dependent variables. Analyses were conducted using the Lavaan package in R version 4.1.3.

7. Results

Two serial mediation models were fitted using maximum likelihood estimators. Bootstrap procedures with 5,000 bootstrap subsamples were conducted for statistical inferences. Figures 2 and 3 visually summarize the results of the serial mediation models. Detailed results of the parameter estimates are presented in Tables 1 and 2.

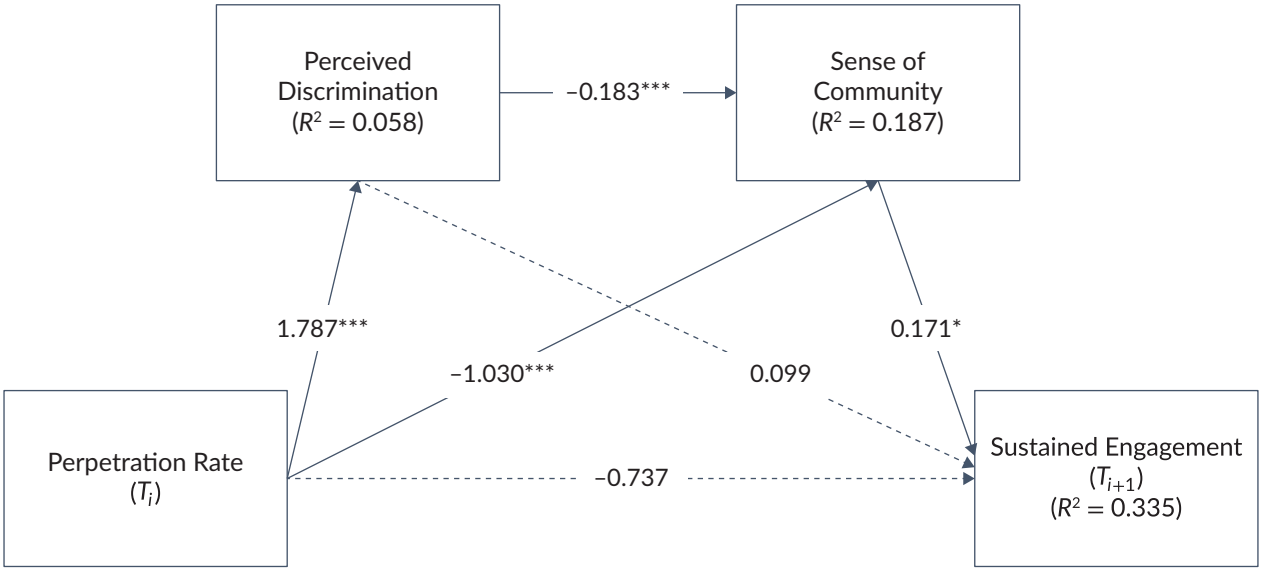


Figure 2. Serial mediation model for perpetration rate on sustained engagement. Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; → significant path; → non-significant path.

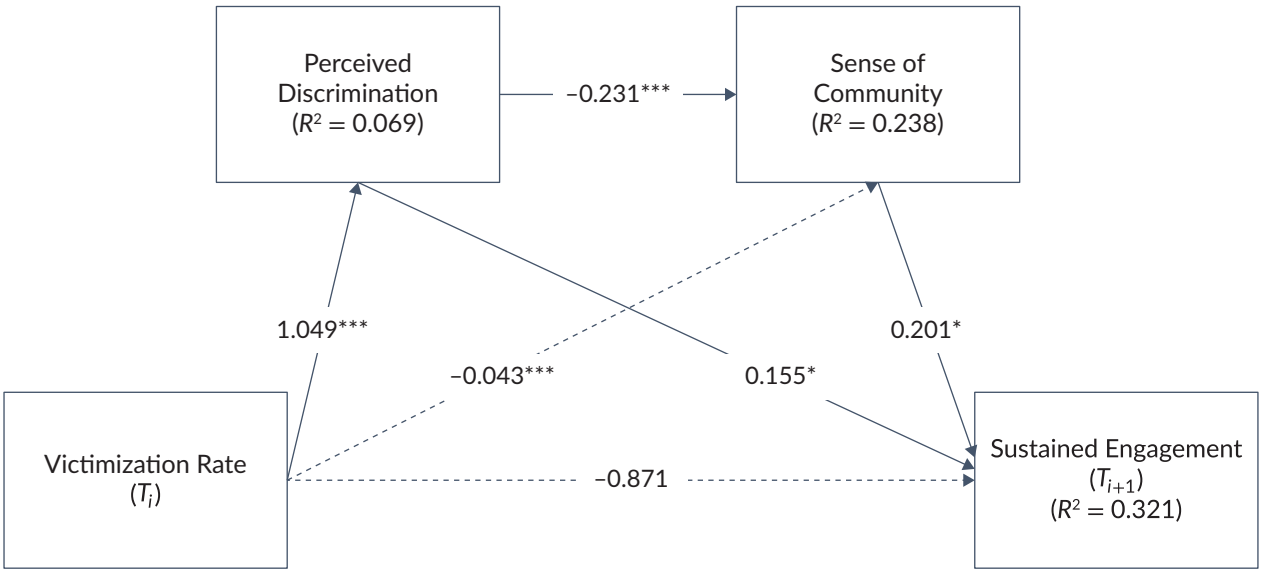


Figure 3. Serial mediation model for victimization rate on sustained engagement. Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; → significance path; → non-significant path.

Table 1. Serial mediation analysis results for perpetration rate on sustained engagement.

Outcome Variable	Predictor Variable	*b*	*SE*	*β*	*p*
Sustained Engagement	Perpetration rate	−0.737	1.257	−0.012	0.557
	Perceived discrimination	0.099	0.060	0.027	0.099
	Sense of community	0.171	0.072	0.044	0.017
	Age	0.325	0.078	0.096	0.000
	Gender	−0.167	0.240	−0.013	0.488
	Race	−0.115	0.061	−0.028	0.060
	Education	0.021	0.058	0.007	0.716
	Income	0.016	0.047	0.007	0.725
	Player rating	0.174	0.065	0.052	0.007
	Socialization motivation	0.057	0.053	0.017	0.276
	Competition motivation	0.036	0.076	0.008	0.631
	Lagged engagement	0.665	0.027	0.550	0.000
Sense of Community	Perpetration rate	−1.030	0.242	−0.063	0.000
	Perceived discrimination	−0.183	0.017	−0.193	0.000
	Age	0.037	0.017	0.042	0.029
	Gender	0.140	0.046	0.042	0.002
	Race	0.000	0.017	0.000	0.988
	Education	−0.088	0.015	−0.110	0.000
	Income	0.008	0.012	0.013	0.505
	Player rating	−0.061	0.014	−0.071	0.000
	Socialization motivation	0.298	0.016	0.351	0.000
	Competition motivation	0.113	0.019	0.098	0.000
Perceived Discrimination	Perpetration rate	1.787	0.382	0.103	0.000
	Age	−0.146	0.019	−0.160	0.000
	Gender	0.041	0.054	0.012	0.455
	Race	0.007	0.019	0.007	0.696
	Education	0.012	0.016	0.014	0.457
	Income	−0.048	0.014	−0.078	0.000
	Player rating	−0.022	0.017	−0.024	0.208
	Socialization motivation	0.063	0.015	0.070	0.000
	Competition motivation	−0.012	0.022	−0.010	0.592
Indirect Effect Parameters					
Perpetration rate → perceived discrimination → sense of community → sustained engagement		−0.056	0.027	−0.001	0.041
Perceived discrimination → sense of community → sustained engagement		−0.031	0.013	−0.008	0.021
Perpetration rate → perceived discrimination → sustained engagement		0.177	0.118	0.003	0.134
Perpetration rate → sense of community → sustained engagement		−0.176	0.084	−0.003	0.035
Total Effect Parameter	Perpetration rate	−0.792	1.255	−0.012	0.528

Table 2. Serial mediation analysis results for victimization rate on sustained engagement.

Outcome Variable	Predictor Variable	*b*	*SE*	*β*	*p*
Sustained Engagement	Victimization rate	−0.871	0.532	−0.034	0.102
	Perceived discrimination	0.155	0.077	0.042	0.045
	Sense of community	0.201	0.088	0.054	0.022
	Age	0.311	0.090	0.096	0.001
	Gender	0.245	0.283	0.021	0.386
	Race	−0.033	0.071	−0.009	0.648
	Education	−0.067	0.069	−0.022	0.334
	Income	0.032	0.055	0.015	0.561
	Player rating	0.146	0.089	0.037	0.102
	Socialization motivation	−0.031	0.069	−0.009	0.654
	Competition motivation	−0.006	0.109	−0.001	0.956
	Lagged Engagement	0.649	0.034	0.535	0.000
Sense of Community	Victimization rate	−0.043	0.132	−0.006	0.746
	Perceived discrimination	−0.231	0.021	−0.235	0.000
	Age	0.002	0.018	0.002	0.924
	Gender	0.117	0.055	0.037	0.033
	Race	−0.035	0.017	−0.035	0.041
	Education	−0.051	0.018	−0.063	0.004
	Income	−0.031	0.014	−0.054	0.025
	Player rating	−0.060	0.020	−0.058	0.003
	Socialization motivation	0.361	0.018	0.407	0.000
	Competition motivation	0.102	0.024	0.078	0.000
Perceived Discrimination	Victimization rate	1.049	0.163	0.151	0.000
	Age	−0.113	0.023	−0.128	0.000
	Gender	−0.068	0.055	−0.021	0.212
	Race	0.000	0.020	0.000	0.982
	Education	0.003	0.020	0.004	0.868
	Income	−0.052	0.015	−0.089	0.001
	Player rating	−0.057	0.022	−0.053	0.009
	Socialization motivation	0.055	0.018	0.061	0.002
	Competition motivation	−0.010	0.027	−0.008	0.717
Indirect Effect Parameters					
Victimization rate → perceived discrimination → sense of community → sustained engagement		−0.049	0.024	−0.002	0.038
Perceived discrimination → sense of community → sustained engagement		−0.046	0.021	−0.013	0.026
Victimization rate → perceived discrimination → sustained engagement		0.162	0.086	0.006	0.059
Victimization rate → sense of community → sustained engagement		−0.009	0.029	0.000	0.767
Total Effect Parameter	Victimization rate	−0.766	0.525	−0.030	0.145

7.1. Perpetration Rate and Sustained Engagement

The analysis revealed that the perpetration rate had a direct effect on one's perceived discrimination ($b = 1.787$, $p < 0.001$, supporting H2) and sense of community ($b = -1.030$, $p < 0.001$), but not on sustained engagement after accounting for its lagged term ($b = -0.737$, $p = 0.557$). Perceived discrimination had a direct effect on sense of community ($b = -0.183$, $p < 0.001$, supporting H3), but not on sustained engagement ($b = 0.099$, $p = 0.099$). Sense of community had a direct effect on sustained engagement ($b = 0.171$, $p = 0.017$, supporting H4). The indirect mediation analysis indicated that the effect of the perpetration rate on sustained engagement was mediated by sense of community ($b = -0.176$, $p = 0.035$), whereas no significant mediation effect was observed through perceived discrimination ($b = 0.177$, $p = 0.134$), supporting H6a. A significant serial mediation was supported whereby perpetration led to increased perceived discrimination, which subsequently decreased the sense of community and, in turn, reduced sustained engagement ($b = -0.056$, $p = 0.041$), supporting H5a. Taken together, the fitted serial mediation model explained approximately 33.5% of the variance in sustained engagement, 18.7% of the variance in sense of community, and 5.8% of the variance in perceived discrimination. A 1% increase in perpetration rate led to a 0.792% decrease in sustained engagement overall. Of this decrease, 0.176% was mediated through a decreased sense of community, and 0.056% was serially mediated through increased perceived discrimination followed by a decreased sense of community.

7.2. Victimization Rate and Sustained Engagement

The results indicated that the victimization rate had a direct effect on one's perceived discrimination ($b = 1.049$, $p < 0.001$, supporting H1), but not on players' sense of community ($b = -0.043$, $p = 0.746$) nor on sustained engagement when controlling for its lagged term ($b = -0.871$, $p = 0.102$). Perceived discrimination exhibited a direct effect on sense of community ($b = -0.231$, $p < 0.001$, supporting H3) and sustained engagement ($b = 0.155$, $p = 0.045$). Sense of community showed a direct effect on sustained engagement ($b = 0.201$, $p = 0.022$), supporting H4. The mediation analysis revealed that the sense of community did not mediate the relationship between victimization rate and sustained engagement ($b = -0.009$, $p = 0.767$), supporting H6b. A significant serial mediation was identified where victimization led to an increase in perceived discrimination, which then led to a decrease in sense of community, culminating in reduced sustained engagement ($b = -0.049$, $p = 0.038$), supporting H5b. Taken together, the fitted serial mediation model explained approximately 32.1% of the variance in sustained engagement, 23.8% of the variance in sense of community, and 6.9% of the variance in perceived discrimination. A 1% increase in victimization rate led to a 0.766% decrease in sustained engagement overall, among which a 0.049% decrease in sustained engagement was serially mediated through increased perceived discrimination and decreased sense of community.

8. Discussion

Theoretically, this study extends the application of RDIM and GEM to online gaming communities, underscoring that perceptions of inclusivity and fairness are central to sustaining active and engaged gaming communities. Incorporating the concept of procedural justice from social psychology offers valuable insights, indicating that individuals value social processes more than game designers and developers might typically recognize. This perspective advocates for a shift in focus from the outcomes of complaint resolution to the dynamics of social interaction. It emphasizes the need for fair procedures, active participation in

decision-making, and the quality of treatment for both reporters and those reported. Practically, this analysis lays out the costs of gaming toxicity in terms of players' sustained engagement in games, thus quantifying the impacts of perpetration and victimization. Highlighting these costs is crucial as it demonstrates the tangible effects of disruptive behaviors on both player engagement and company revenue. For instance, even a decrease of one battle count can lead to a reduction of approximately five to 20 minutes of a player's time in games. This not only represents a significant opportunity cost in potential revenue lost for gaming companies but also lessens the quality of time players spend engaging in meaningful gameplay. In addition, the results suggest that sense of community is the key to keeping players engaged. In response, gaming platforms and community managers are encouraged to develop strategies that minimize players' perceptions of discrimination and provide effective mechanisms for voicing concerns or disputing reports, thus maintaining players' sense of community and fostering a safer and more inclusive gaming environment.

In addition to the serial mediation model—the common pathway—this study also identified unique mechanisms for perpetrators and victims. For perpetrators, the pathway to disengagement is mediated by a decreased sense of community, rather than directly by feelings of discrimination. This highlights a nuanced aspect of the perpetrator's experience; although they might feel discriminated against after being reported, it is primarily the erosion of their sense of belonging that drives their disengagement. This observation provides a practical insight for gaming community management: Maintaining an inclusive community that offers support and integration opportunities for all players, including those who have been reported, might reduce disengagement rates. Furthermore, reported players should be respectfully given a voice, such as the opportunity to dispute or explain themselves, demonstrating that they are valued members of the community unless proven disruptive.

As predicted by the GEM (Tyler & Blader, 2003), the sense of community among victims remains intact, likely due to the reporting system that allows them to voice concerns rather than disengage directly. This underscores the importance of procedural justice in gaming environments, suggesting that the manner in which toxicity issues are handled is as crucial as the outcome itself. For victims, time in-game is only reduced if their perceived discrimination leads to a decreased sense of community. Otherwise, even if they feel discriminated against, it does not translate into decreased play. This finding highlights an interesting observation: While reporting and being reported has a cost in terms of players' time spent in games, not having the reporting system may incur a greater cost, as it provides everyone with an alternative pathway to voice their experiences before deciding to withdraw from the game. In addition, according to the GEM, procedures that promote group participation are perceived as fairer, regardless of which party receives support from the gaming community. In some games, involving community members in making crowdsourced judgments about whether a reported toxic player should be punished can enhance fairness (Blackburn & Kwak, 2014). Although we cannot verify this with the data here, GEM suggests that both parties involved are likely to view this participatory process as fair and just, even if the final outcome does not favor them. Additionally, games might consider implementing features that unite players by recognizing their commonalities rather than highlighting features that divide players, especially in competitive games.

Underpinning RDIM and GEM is the SIT, which emphasizes the importance of social identity and the processes of social categorization, identification, and comparison in online video games (Trepte & Loy, 2017). The fundamental reason for individuals' engagement or disengagement from the gaming community hinges on their self-perception and their connection with the community. Consequently, games should aim to enhance players' self-esteem, not through traditional social comparison processes but through the benefits

that the gaming community provides, such as a sense of competence from collaboration and a sense of relatedness from socialization and cooperation (Ryan et al., 2006). Efforts should be made to minimize cues that divide players, such as overt attention to differences in player skills or other indicators of offline demographic identities.

While this study primarily addresses the repercussions of toxic behavior rather than its origins, it acknowledges the potential for a reciprocal relationship and role fluidity within the toxicity cycle (Kordyaka et al., 2023). SIT suggests that within computer-mediated environments, individuals' self-identity may be dominated by their social identity and align their behavior with prevailing group norms. Experiencing victimization or exposure to gaming toxicity can alter an individual's perception of these norms (Shen et al., 2020) and provoke frustrations that necessitate coping with further toxic behavior (Neto et al., 2017), potentially perpetuating a cycle of toxicity. Moreover, the subjective nature of toxic experiences, particularly in ambiguous situations, may lead to inconsistencies between the punishments developers impose and the players' perceptions of toxicity, as well as disparities between what is universally recognized as toxic and what is considered acceptable by perpetrators. The reporting system may exacerbate feelings of unfairness among those penalized, which is also a loss for gaming communities and companies. A thorough understanding of toxicity's consequences is incomplete without an exploration of its developmental dynamics. Future research may compare self-reported and behaviorally reported data on toxicity to enhance understanding of the fluid dynamics among players within the cycle of toxicity, the subjectivity of these experiences, and the interplay between the causes and consequences of gaming toxicity.

The study is not without its limitations, which future research could help address. Firstly, while the study employs a longitudinal design, it does not establish robust causal links between involvement in gaming toxicity, discrimination, sense of community, and engagement. Future research could employ laboratory or field experiments to clarify these causal relationships. Field experiments might be preferable, as previous research on procedural justice suggests that effects are generally stronger in field settings than in laboratory settings (Lind & Tyler, 1988). Secondly, the study's focus on WoT—a competitive-cooperative online multiplayer game with unique demographics—limits its empirical generalizability to similar games. Further research is needed across different genres and types of games with diverse player demographics.

Acknowledgments

We thank Wargaming for providing access to their anonymized data, and Eugene Kislyi and Jeremy Ballenger for their valuable feedback.

Conflict of Interests

The third author was engaged as a consultant for the company that provided the study's data. Work on this research project was not compensated.

Data Availability

Data will be made available on request with permission of Wargaming.

References

Adachi, P. J., Hodson, G., & Hoffarth, M. R. (2015). Video game play and intergroup relations: Real world implications for prejudice and discrimination. *Aggression and Violent Behavior*, *25*, 227–236.

Anti-Defamation League. (2023). *Hate in no game: Hate and harassment in online games 2023*. https://www.adl.org/resources/report/hate-no-game-hate-and-harassment-online-games-2023

Blackburn, J., & Kwak, H. (2014). STFU NOOB! Predicting crowdsourced decisions on toxic behavior in online games. In C.-W. Chung (Ed.), *WWW '14: Proceedings of the 23rd international conference on World wide web* (pp. 877–888). ACM.

Gray, K. L. (2012). Deviant bodies, stigmatized identities, and racist acts: Examining the experiences of African-American gamers in Xbox Live. *New Review of Hypermedia and Multimedia*, *18*(4), 261–276.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard University Press.

Jasinskaja-Lahti, I., Celikkol, G., Renvik, T. A., Eskelinen, V., Vetik, R., & Sam, D. L. (2018). When psychological contract is violated: Revisiting the rejection-disidentification model of immigrant integration. *Journal of Social and Political Psychology*, *6*(2), 484–510.

Jasinskaja-Lahti, I., Liebkind, K., & Solheim, E. (2009). To identify or not to identify? National disidentification as an alternative reaction to perceived ethnic discrimination. *Applied Psychology: An International Review*, *58*(1), 105–128. https://doi.org/10.1111/j.1464-0597.2008.00384.x

Kahn, A. S., Shen, C., Lu, L., Ratan, R. A., Coary, S., Hou, J., Meng, J., Osborn, J., & Williams, D. (2015). The trojan player typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations. *Computers in Human Behavior*, *49*, 354–361. https://doi.org/10.1016/j.chb.2015.03.018

Kim, J. (2011). Developing an instrument to measure social presence in distance higher education. *British Journal of Educational Technology*, *42*(5), 763–777.

Kordyaka, B., Jahn, K., & Niehaves, B. (2020). Towards a unified theory of toxic behavior in video games. *Internet Research*, *30*(4), 1081–1102.

Kordyaka, B., Laato, S., Jahn, K., Hamari, J., & Niehaves, B. (2023). The cycle of toxicity: Exploring relationships between personality and player roles in toxic behavior in multiplayer online battle arena games. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CHI PLAY), Article 397. https://doi.org/10.1145/3611043

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer; Business Media.

Neto, J. A., Yokoyama, K. M., & Becker, K. (2017). Studying toxic behavior influence and player chat in an online video game. In A. Sheth (Ed.), *WI '17: Proceedings of the International Conference on Web Intelligence* (pp. 26–33). ACM. https://doi.org/10.1145/3106426.3106452

Nguyen, S. H., Sun, Q., & Williams, D. (2022). How do we make the virtual world a better place? Social discrimination in online gaming, sense of community, and well-being. *Telematics and Informatics*, *66*, Article 101747. https://doi.org/10.1016/j.tele.2021.101747

Paradies, Y. (2006). Defining, conceptualizing and characterizing racism in health research. *Critical Public Health*, *16*, 143–157.

Passmore, C. J., & Mandryk, R. L. (2020). A taxonomy of coping strategies and discriminatory stressors in digital gaming. *Frontiers in Computer Science*, *2*, 40.

Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022). Feeling good and in control: In-game tools to support targets of toxicity. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CHI PLAY), Article 235.

Rovai, A. P. (2002). Building sense of community at a distance. *International Review of Research in Open and Distributed Learning*, *3*(1), 1–16.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360.

Sarason, S. B. (1974). *The psychological sense of community*. Jossey-Bass.

Shen, C., Sun, Q., Kim, T., Wolff, G., Ratan, R., & Williams, D. (2020). Viral vitriol: Predictors and contagion of online toxicity in World of Tanks. *Computers in Human Behavior*, *108*, Article 106343.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, *56*(65), Article 9780203505984–16.

Trepte, S., & Loy, L. S. (2017). Social identity theory and self-categorization theory. In P. Rössler (Ed.), *The international encyclopedia of media effects*. https://doi.org/10.1002/9781118783764.wbieme0088

Tseng, F.-C., Huang, H.-C., & Teng, C.-I. (2015). How do online game communities retain gamers? Social presence and social capital perspectives. *Journal of Computer-Mediated Communication*, *20*(6), 601–614.

Tseng, F.-C., & Teng, C.-I. (2014). Antecedents for user intention to adopt another auction site. *Internet Research*, *24*(2), 205–222.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Blackwell.

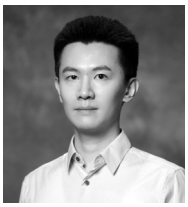
Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, *7*(4), 349–361.

Williams, D. R., Yu, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *Journal of Health Psychology*, *2*(3), 335–351.

Zagenczyk, T. J., Cruz, K. S., Woodard, A. M., Walker, J. C., Few, W. T., Kiazad, K., & Raja, M. (2013). The moderating effect of Machiavellianism on the psychological contract breach: Organizational identification/disidentification relationships. *Journal of Business and Psychology*, *28*, 287–299.

About the Authors

Mingxuan Liu (PhD) is an assistant professor in the Department of Communication at the University of Macau. Her research explores how emerging technologies shape user behavior, health, and well-being, and vice versa. She also makes causal inferences on user behavior change using observational data, panel surveys, natural experiments, and psychophysiological experiments. Her work has appeared in peer-reviewed journals, including *New Media & Society*, *Media Psychology*, *Health Communication*, *Computers in Human Behavior*, and *Mobile Media & Communication*.



Jack Lipei Tang (PhD) is an assistant professor in the Department of Advertising and Public Relations at The University of Alabama. His research focuses on advocacy communication, social media, social networks, and computational social science. His work has been published in various communication, information science, and interdisciplinary journals, including *Computers in Human Behavior*, *Social Media + Society*, *Information Society*, and *Social Networks*. He earned his PhD from the Annenberg School for Communication and Journalism at the University of Southern California.

Dmitri Williams is a professor at the University of Southern California Annenberg School for Communication. He studies games, technology, and community.