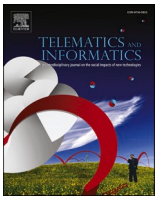
Telematics and Informatics 106 (2026) 102382

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele



The empowerment effect of jury-based content moderation: Impact on user stickiness on social media platforms

Baojun Ma, Xiaoyan Wen, Yi Chen *, Yao Mu, Yi Huang

Key Laboratory of Brain-Machine Intelligence for Information Behavior (Ministry of Education and Shanghai), School of Business and Management, Shanghai International Studies University, Shanghai 201620, PR China

ARTICLE INFO

Keywords:
Jury-based content moderation
Platform governance
User generated content
Social media
User stickiness
Social identity

ABSTRACT

The increasing user-generated content on social media platforms, while enriching user experiences, has posed higher risks of exposure to inappropriate content, which can diminish user satisfaction and platform stickiness—a critical factor of platform success. Jury-based content moderation has emerged as a critical community-driven governance approach that empowers users to collectively evaluate the appropriateness of user-generated content. This study investigates whether and how granting users jury-based content moderation power influences their stickiness on social media platforms. Through three experiments, the research demonstrates a positive impact of jury-based content moderation power on user stickiness, mediated by users' social identity. This effect can be further amplified by reputation rewards and power scarcity. Moreover, through a quasi-experimental study based on observational data from a leading Chinese online video platform, we further confirm the positive effect of jury-based content moderation, validating generalizability of our findings in real-world context. By uncovering the linkage and boundary conditions between jury-based content moderation and platform stickiness, this study not only enriches our understanding on the empowerment effect of participatory governance but also provides actionable strategies for leveraging jury-based moderation and associated incentive schemas to foster user retention.

1. Introduction

The various forms of user-generated content on social media platforms, such as comments and danmaku, have become integral to platform users' content consumption experience (Wang, 2023). While valuable user-generated content may improve user experience and engagement (Wei, 2023), inappropriate content, such as hate speech and cyberbullying, may pose risks to user satisfaction (Wang, 2021; Ejaz et al., 2024; Wachs et al., 2024). This challenge is particularly salient in social media platforms as content consumption is deeply intertwined with social interaction. Harmful user-generated content not only imposes negative emotions to individual users but also disrupts the broader community atmosphere, ultimately threatening users' platform stickiness – defined as users' dependency and loyalty to a platform, which is critical to platforms' long-term competitive advantage and sustainability (Rong et al., 2019). High user stickiness could facilitate continued platform usage and encourage user interactions (Pang and Zhang, 2024); whereas declining stickiness may lead to user attrition and destabilize the platform's ecosystem (Hsu et al., 2017).

* Corresponding author at: Block 2-201, 1550 Wenxiang Road, Shanghai 201620, China.
E-mail addresses: mabaojun2003@163.com (B. Ma), 0214101695@shisu.edu.cn (X. Wen), cheny@shisu.edu.cn (Y. Chen), muy@shisu.edu.cn (Y. Mu), 2019043@shisu.edu.cn (Y. Huang).

https://doi.org/10.1016/j.tele.2026.102382
Received 12 December 2024; Received in revised form 14 September 2025; Accepted 16 February 2026
Available online 17 February 2026

To combat inappropriate content, platforms started to rely on their user community for moderation (Seering, 2020). This has given rise to a novel approach called jury-based content moderation, a crowdsourced mechanism that empowers platform users to serve as digital jurors (Zhao and Hobbs, 2025). Unlike traditional content governance measures, such as professional auditors and algorithmic detection tools (Spence et al., 2023; Gorwa et al., 2020), jury-based content moderation is both economical (Hettiachchi and Goncalves, 2020) and community-driven (Zhao and Hu, 2025). Many major platforms have already implemented this governance approach. For example, Weibo uses a crowdsourced user "committee" system to empower users to make content moderation decisions, while Bilibili has established user-based disciplinary committees responsible for examining user complaints and determining appropriate actions.

The effectiveness of jury-based content moderation or similar forms of crowdsourced content moderation has been widely examined in prior research (Micek and Solovey, 2024; Fan and Zhang, 2020). For instance, Drolsbach et al. (Drolsbach et al., 2024) found that crowd-sourced "community notes" significantly increased users' trust in fact-checks and their accuracy in identifying misinformation. Fan and Zhang (Fan and Zhang, 2020) found that digital juries significantly enhance users' perceptions of procedural justice within moderation systems. Hu et al. (Hu et al., 2021) showed that decisions made by online juries are often perceived as more legitimate than those made by algorithms. In addition, jury-based moderation has also been shown to influence user behavior. Chuai et al. (Chuai et al., 2024) examined whether crowd-sourced "community notes" could reduce user engagement (i.e., behaviors of retweets and likes) with misleading posts, while Zhao and Hobbs (Zhao and Hobbs, 2025) demonstrated that public sanctions issued by user juries for personal attacks led to a short-term reduction in offensive behavior among the reported users.

However, existing studies tend to understand jury-based content moderation from the perspective of users being moderated, examining how it shapes users' attitudes toward evaluative decisions or affects users' online behaviors. Less attention has been paid to individuals who serve as digital jurors and how the content moderation experience influences their platform stickiness. On social media platforms, users may play a tripartite role: they could be content consumers (potential recipients of inappropriate user-generated content), content producers (potential sources of disputed content), and active participants in jury-based content moderation. While the first two roles expose users to the consequences of moderation, the third role positions them as active contributors to the enforcement of platform norms. This unique positioning—being both governed and governor of content moderation—underscores the necessity of studying jury-based content moderation from the juror's perspective. The experience of participating in jury-based content moderation, whether positive or negative, might significantly impact their future engagement and activeness on the platform.

Meanwhile, jury-based content moderation also differs fundamentally from traditional methods used to increase user stickiness. Previous research has largely focused on platform-driven approaches, such as improving user interfaces, providing personalized content and services (Li et al., 2024; Li et al., 2021); and introducing interactive features (Pang et al., 2024). While these findings reveal the importance of useful functionality and hedonic value in user retention, they might overlook the role of users' subjective agency. Jury-based content moderation, in contrast, engages users as managers of the platform's social and normative environment, shifting users from passive audience to active governors. Thus, by investigating the psychological perceptions of these empowered users, we may gain new insights into improving social media users' platform stickiness that goes beyond conventional design strategies.

Given these gaps and opportunities, our study aims to address a critical question: how does jury-based content moderation affects social media users' platform stickiness and what are the underlying mechanisms? According to social identity theory, users who perceive alignment with the community's values and goals are more likely to develop a sense of social identity and belonging, thus being more inclined to contribute to the community (Gao et al., 2022). Following this logic, this study employs social identity as a framework to assess the role of jury-based content moderation in sustaining user activity and commitment to social media communities.

Furthermore, our study also attempts to explore potential moderating factors that may adjust the relationship between jury-based content moderation and user stickiness. Guided by self-determination theory (Deci et al., 2017), we choose to investigate moderating effects of both internal and external incentives. Specifically, we focus on two key moderators: reputation rewards and power scarcity. Reputation rewards are a typical type of external incentive that provides social recognition and status, which may potentially affect users' social identification upon gaining jury-based content moderation power. Power scarcity, on the other hand, acts as an internal incentive derived from social comparison (Taylor and Lobel, 1989). It can satisfy users' intrinsic needs (e.g., achievement and meaningfulness), thus enhancing the perceived value of jury-based content moderation opportunities (Park et al., 2022). By examining these moderators, our study provides insights into optimizing incentives and exclusivity to enhance empowerment of jury-based content moderation power.

To explore the proposed research questions, our study uses danmaku video websites as the research setting. Danmaku (or "barrage") is a novel commenting function where users post comments that appear at specific times during videos, facilitating "quasi-synchronous" interactions with the video content and other viewers (Wei, 2023). Widely adopted across major Chinese video platforms, danmaku's high volume and immediacy present significant content management challenges that may disrupt viewing experience (Wang, 2021). In response, platforms like Bilibili empower users to evaluate inappropriate danmaku, making this context particularly relevant for studying jury-based content moderation effects. Through three controlled experiments and a quasi-experimental using real-world data from Bilibili's "Disciplinary Committee Weekly Reports", our study empirically explores users' psychological states and user stickiness following participation in jury-based content moderation of danmaku, and further investigate the behavioral effects of being publicly commended in the weekly reports, focusing on how such recognition shapes users' continued engagement. The results reveal that jury-based content moderation significantly enhances users' social identity with the platform, thus increasing their platform stickiness. In addition, this effect can be amplified by both reputation rewards and power scarcity. The quasi-experimental analysis additionally shows that public commendation leads to an active engagement.

This study makes several key contributions. First, our study shifts the focus from users being moderated to those acting as digital

jurors, highlighting that jury-based content moderation can function not only as a governance mechanism—shaping user attitude and behavior within the platform—but also as a strategy to engage and retain users. Second, by examining jury-based content moderation through a lens of user empowerment, we highlight the potential of user's subjective agency in platform governance, extending our understanding of user stickiness enhancement beyond traditional platform-driven approaches. Third, our study identifies social identity as a key mechanism though which jury-based content moderation influences user stickiness, while demonstrating how reputation rewards and power scarcity serve as important boundary conditions that could further enhance the impact of jury-based content moderation. It provides an integral framework combining both social identity theory and self-determination theory to understand the effect of jury-based moderation on user stickiness. Finally, our study provides practical guidance for platform managers seeking to leverage jury-based content moderation to foster a stronger sense of user identification and enhance user stickiness. It suggests the need to incorporate incentive strategies into the design and optimization of jury-based content moderation practice.

2. Theory and hypotheses

2.1. Jury-based content moderation, social identity, and user stickiness

Jury-based content moderation, as a type of self-moderation approach, represents a fundamental shift in content management authority, transferring the power from platform operators to users (Zhao and Hobbs, 2025). It endows users with supervisory power and decision-making responsibility to assess content compliance with platform rules. In social media contexts, jury-based content moderation could serve as a form of user empowerment that involves users more deeply in platform governance. Existing research demonstrates that empowering users could create positive emotional experiences and encourage users to participate in interpersonal interactions (Werner et al., 2022). For instance, Mostafa and Sobhy Temerak (Mostafa and Sobhy Temerak, 2024) found that consumer empowerment positively influences Facebook brand page stickiness by shaping users' positive experiences. Similarly, Ali Acar and Puntoni (Acar and Puntoni, 2016) found that brands could enhance customer loyalty through customer empowerment measures, such as inviting consumers to contribute to content creation and evaluation or to improve others' ideas. In sum, by creating a perception of control, jury-based content moderation could increase users' emotional connection to the platform, thereby enhancing their platform stickiness. Thus, we hypothesize:

H1: Jury-based content moderation positively influences user stickiness of the platform.

Building on the previous arguments, we further propose that social identity is a key psychological mechanism underlying the impact of jury-based moderation. According to social identity theory, individuals who perceive their values and goals align with a community are more likely to develop a sense of belonging and self-identification, which in turn increases their willingness to contribute (Li et al., 2021). In social media context, jury-based content moderation enables its participants to express their personal opinions on community standards. By empowering users to affect platform operations, this approach elevates users from ordinary community members to community managers and system builders (Lampe et al., 2014). This role transformation provides opportunities for users to align their efforts with the platform's goals (Preece and Shneiderman, 2009). When users feel valued and aligned with community norm, they may perceive higher community integration and social identification, thus motivating them to invest more time and resources in maintaining and promoting community (Ren et al., 2007).

Existing studies show that heightened perception of social identity is a crucial driver of sustained community engagement (Zhang and Liu, 2024; Xu et al., 2025). Specifically, social identification could enhance sense of affiliation and self-expression, which motivates users to actively participate in community activities (Bagozzi and Dholakia, 2006; Shams et al., 2024). When users feel they belong and are an integral part of the community, they are more willing to invest time and resources in maintaining it (Ren et al., 2007). In sum, jury-based content moderation could help foster users' social identity and create a deeper user-platform attachment, thus resulting in improved user retention and stickiness. Thus, we hypothesize:

H2: Social identity mediates the positive effect of jury-based content moderation on user stickiness of the platform.

2.2. Moderating effect of reputation rewards

While users' identification mediates the relationship between jury-based content moderation and stickiness, its formation does not solely originate from the jury-based content moderation power itself but also depends on the value obtained from jury-based moderation participation (Dholakia et al., 2004). Jury-based content moderation may lead to stronger identification if it offers users more value (Chen and Tsai, 2020).

Perceived value can be both extrinsic and intrinsic. On social media platforms, extrinsic values often include non-monetary benefits like reputation, social status, and personal image (Zhang et al., 2023). Research has demonstrated that reputation rewards increase perception of social value, particularly by elevating users' community status through acknowledging their contributions (Wasko and Faraj, 2005). The increased social value can fulfill users' needs to improve self-image and gain others' recognition and respect (Singh et al., 2021), thereby strengthening their sense of community belonging and social identity.

In sum, reputation rewards received from jury-based content moderation participation not only satisfies users' needs for social status and favorable self-image (Mou et al., 2025) but also signal the community's recognition of their contributions (Elsharnouby et al., 2021). This could increase users' perception of social value, which ultimately enhances user stickiness (Mou et al., 2025). Under this circumstance, jury-based content moderation is more than just content moderation power—it becomes a source of social value that boosts users' connection to and identification with the community. Thus, we hypothesize:

H3a: Reputation rewards positively moderate the effect of jury-based content moderation on social identity, such that the effect

becomes more positive when users receive reputation rewards compared to when they do not.

H3b: Reputation rewards positively moderate the indirect effect of jury-based content moderation on stickiness via social identity, such that the indirect effect becomes more positive when users receive reputation rewards compared to when they do not.

2.3. Moderating effect of power scarcity

In social media, intrinsic value of jury-based content moderation typically comes from satisfaction users gain through conducting the judgements. This includes emotional values like self-worth and self-actualization (Sun et al., 2012). Power scarcity of jury-based content moderation refers to the range of users eligible to participate in jury-based moderation. According to social identity theory, individuals tend to distinguish themselves from others in social environments. Perceptions of identity distinctiveness can trigger wishful identification among the audience (Lin and Huang, 2024). Higher power scarcity implies fewer selected participants, which amplifies the relative influence and importance of each participant (Weinstein, 2022; Shi et al., 2024). By limiting participants' size, users feel their role within the community is crucial, viewing their judgment as more than just duties but as expressions of their capabilities. This recognition boosts self-esteem and meets their psychological needs for achievement (Deci and Ryan, 2013). Prior research indicates that when a group identity satisfies one's self-enhancement needs, it becomes more attractive, further strengthening one's identification with the group and encouraging in-group cooperation (Gu et al., 2022). Similarly, Grant (Grant, 2008) emphasizes that perceiving one's actions as socially impactful fosters a stronger identification with the community. Therefore, higher power scarcity of jury-based content moderation may lead to greater appreciation for being selected as judges, which strengthens social identity and ultimately enhances user stickiness.

Moreover, the theory of commodity scarcity suggests that scarce resources are deemed more valuable and desirable due to their perceived uniqueness and irreplaceability, driving greater pursuit and appreciation (Li et al., 2024). Tajfel and Turner (Tajfel and Turner, 1979) argue that uniqueness increases group identification by distinguishing the group from others. Therefore, limiting the number of jury-based content moderation participants can make users view their roles as more unique, which further strengthens their social identity perception. Thus, we hypothesize:

H4a: Power scarcity positively moderates the effect of jury-based content moderation on social identity, such that the effect becomes more positive when power scarcity of jury-based content moderation is high compared to when it is low.

H4b: Power scarcity positively moderates the indirect effect of jury-based content moderation on stickiness via social identity, such that the indirect effect becomes more positive when power scarcity of jury-based content moderation is high compared to when it is low.

The research model of this study is summarized in Fig. 1.

3. Experimental design and data analysis

To investigate the impact of jury-based content moderation on user stickiness of the video platform, we designed three online user experiments. We recruited participants with varied demographics from the Credamo platform, a popular Chinese online survey platform that offers access to a large and diverse participant pool. To maintain the quality of survey responses, we implemented several screening criteria during sample collection, such as attention-check items and sample quality filtering. The experiments simulated real-world danmaku viewing and jury-based content moderation using the interface of Bilibili, China's leading danmaku video platform but removed the platform's logo from screenshots to reduce bias from user familiarity.

3.1. Experiment 1: Effect of jury-based content moderation on user stickiness and mediating effect of social identity

Experiment 1 was conducted to test the main effect of jury-based content moderation on user stickiness and the mediating effect of social identity (i.e., H1 and H2), using a single-factor between-groups design. Sample size estimation using G*Power (effect size $f = 0.5$, $\alpha = 0.05$, power $= 0.8$) indicated that 102 participants were minimum required for independent-sample *t*-test (Faul et al., 2007).

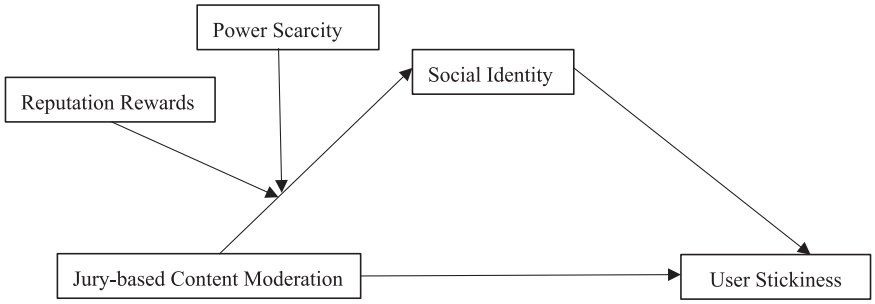


Fig. 1. Theoretical model.

3.1.1. Stimulus materials and procedure

The experiment collected 20 danmaku comments from different channels on Bilibili, recruited 130 subjects, and randomly divided them into two groups: with and without jury-based content moderation power. Informed consent was obtained before the experiment began. Participants received a brief overview of the goals, privacy protection measures, experimental procedures, and duration. To immerse subjects in the experimental scenario, both groups first read the following materials:

Danmaku, on-screen comments that appear alongside videos, have become increasingly popular. However, this rise in popularity has also led to uncertain quality of danmaku content, which can disrupt the user experience and negatively impact video platforms.

Next, to manipulate their perception of jury-based content moderation power, the two groups were given different instructions. The with-power group was informed they were invited as jurors to participate in danmaku governance on an online video platform, while the without-power group was told they were invited as investigators for a market research agency to evaluate danmaku quality. Both groups were shown 20 video screenshots, each highlighting a danmaku comment, and asked to vote on whether each comment was appropriate. Low-quality comments such as messages violating laws and regulations, spam advertisements, and vulgar/ridicule/provocation words are common instances that might be voted as improper danmaku. After voting, the two groups received different feedback. The with-power group was informed that: *Your votes have been received by the video platform! They will be used for platform's final adjudication strategy.* The without-power group was informed that: *Your votes have been received by the research agency! They will be used only for research purposes.*

After voting and feedback, subjects completed several scales measuring user stickiness (Li et al., 2024; Huang and Chung, 2024) (4 items, e.g., "I intend to continue using this online video platform instead of other alternative services or platforms", Cronbach's $\alpha = 0.790$), social identity (Bagozzi and Dholakia, 2006; Farivar and Wang, 2022) (10 items, e.g., "I feel a sense of belonging to this online video platform", Cronbach's $\alpha = 0.891$), and perceived control (Huh et al., 2023; Chun and Lee, 2017) over jury-based content moderation power (4 items, e.g., "During the danmaku display, my vote would be accepted by the video platform"). All measurement items were derived from previous studies and modified to suit the context of danmaku video platform (see Appendix A for complete scale). All variables were assessed using a 7-point Likert scale from 1 ('strongly disagree') to 7 ('strongly agree'). Finally, demographic information was gathered. Each participant received 5 RMB upon completion of the experiment.

3.1.2. Results

To ensure data validity, this study excluded samples that failed attention checks ($n = 4$, 3.08%) and one participant (0.77%) who exhibited patterned responses. Finally, 125 valid questionnaires (76 female; majority aged 21–40; most holding an undergraduate degree) were collected, achieving a 96.15% effective response rate. Common method bias was tested through Harman's single factor test: an exploratory factor analysis (EFA) in SPSS did not return any single loading factor accounting for more than 50% of the variance (Podsakoff et al., 2003).

3.1.2.1. Manipulation Check. To test the effectiveness of the jury-based content moderation power manipulation, an independent sample T-test analysis is conducted. Results indicated a significant difference in perceived control between subjects with jury-based content moderation power ($M = 6.18$) and those without ($M = 5.26$), with $p < 0.001$, confirming successful manipulation of jury-based content moderation power.

3.1.2.2. Main effect. To verify the influence of jury-based content moderation power on user stickiness, an independent sample T-test was conducted. Results showed a significant difference in user stickiness between the two groups ($M_{with-power} = 5.83$, $SD = 0.63$; $M_{without-power} = 5.31$, $SD = 1.00$; $p = 0.001$). This confirms that jury-based content moderation significantly enhances user stickiness of the platform, supporting H1.

3.1.2.3. Mediating effect of social identity. PROCESS Model 4 was used with the bootstrap method with 5000 samples and a 95% confidence interval to confirm the mediating relationships (Hayes, 2009). As shown in Table 1, the results indicate that the mediating effect of social identity is significant at the 95% confidence interval (LLCI = 0.19, ULCI = 0.64, not including 0), with an effect size of

Table 1
Coefficients of the mediation effect on social identity.

	Social Identity (Mediator)			User Stickiness (Dependent Variable)		
	Coeff.	SE	p	Coeff.	SE	p
Constant	4.81	0.21	0.000	1.08	0.41	0.009**
Jury-based Moderation Power	0.51	0.13	0.000***	0.12	0.12	0.301
Social Identity				0.77	0.08	0.000***
			Effect	SE	LLCI	ULCI
Total Effect			0.51	0.15	0.22	0.81
Direct Effect			0.12	0.12	−0.11	0.35
Indirect Effect			0.39	0.12	0.19	0.64
	$R^2 = 0.11$			$R^2 = 0.50$		
	$F = 15.18$, $p = 0.000$			$F = 61.86$, $p = 0.000$		

Notes: *p < 0.05, **p < 0.01, ***p < 0.001.

0.39. After controlling for social identity, the direct effect of jury-based content moderation power on user stickiness is not significant (LLCI = -0.11, ULCI = 0.35, including 0). This demonstrates that the mediating effect of social identity is significant, thus supporting H2.

3.2. Experiment 2: Moderating effect of reputation rewards

To verify the moderating effect of reputation rewards on the association between jury-based content moderation and social identity and the moderated mediation effect on user stickiness via social identity (i.e., H3a and H3b), the second experiment was conducted using a 2 (with/without jury-based content moderation power) × 2 (with/without reputation rewards) between-groups design. Sample size estimation for the F-test using G*Power (effect size f = 0.25, α = 0.05, number of groups = 4, power = 0.8) indicated a minimum of 179 participants (Faul et al., 2007).

3.2.1. Procedure

The experiment recruited 240 subjects and randomly divided them into four groups: with-power-with-rewards, with-power-without-rewards, without-power-with-rewards, and without-power-without-rewards, while ensuring no subjects had participated in Experiment 1. Following the main process of Experiment 1, subjects voted on 20 danmaku comments, with the with-power and without-power groups receiving different manipulations in identity and voting feedback. To differentiate the groups with and without reputation rewards, subjects in with-rewards groups were informed they would receive an honorary certificate from the video platform/research agency after voting, while the without-rewards groups received no such information. For with-rewards groups, an electronic honorary certificate was presented to each subject upon completion of voting.

Finally, subjects completed scales measuring user stickiness (Cronbach's α = 0.835), social identity (Cronbach's α = 0.897), and perceived control over jury-based content moderation power, and completed the manipulation check of reputation rewards (McKernan et al., 2015) (e.g., “After participating in the danmaku quality survey/governance, I received an honorary reward from the research agency/video platform. ”). All scales were based on a 7-point Likert scale. Subjects' demographic information was collected. Each participant received 7 RMB upon completion.

3.2.2. Results

To ensure data quality, this study excluded participants who failed attention checks (n = 18, 7.50%), had excessively brief response times (n = 3, 1.25%), or showed patterned responses (n = 2, 0.83%). Ultimately, 217 valid questionnaires (109 female; majority aged 21–40; most holding an undergraduate degree) were collected, achieving a 90.42% recovery rate. A Harman's single-factor test was conducted as in Study 1, confirming no significant common method bias.

The independent samples T-test results confirmed successful manipulation of perceived control over jury-based content moderation power and reward. Participants with jury-based content moderation power had a significantly higher perceived control (M = 5.98) compared to those without (M = 4.90), with $p < 0.001$. Reward group participants scored significantly higher in reward perception (M = 6.66) than those in the non-reward group (M = 3.80), with $p < 0.001$. The results of two-way ANOVA (Table 2) indicate that participants with jury-based content moderation power had a significantly higher user stickiness compared to those without power ($M_{with-power} = 5.86$, SD = 0.83; $M_{without-power} = 5.30$, SD = 1.08, $p < 0.001$), consistent with the findings of Experiment 1.

3.2.2.1. Moderating effect of reputation rewards. The results of two-way ANOVA (Table 2) indicate that the interaction between jury-based content moderation power and rewards significantly affect social identity ($p < 0.05$), confirming the moderating role of rewards in the influence of jury-based content moderation power on users' social identity. To reveal how rewards moderate the effect of jury-based content moderation power on social identity, we conducted a simple slope test and drew a simple effect analysis plot based on

Table 2
Two-way ANOVA results of moderated by reputation rewards.

		User Stickiness			Social Identity		
X	W	Mean	SD	p	Mean	SD	p
Without Power	Without Reward	5.32	0.85	0.846	5.27	0.83	0.888
	With Reward	5.28	1.29		5.29	1.07	
	Total	5.30	1.08		5.28	0.95	
With Power	Without Reward	5.59	1.00	0.004**	5.61	0.72	0.002**
	With Reward	6.13	0.48		6.01	0.38	
	Total	5.86	0.83		5.84	0.62	
Total	Without Reward	5.45	0.93	0.054	5.44	0.79	0.021*
	With Reward	5.69	1.07		5.68	0.90	
	Without Power	5.30	1.08	0.000***	5.28	0.95	0.000***
	With Power	5.86	0.83		5.84	0.62	
X				0.000***			0.000***
W				0.054			0.021*
X * W				0.028*			0.034*

Notes:*p < 0.05, **p < 0.01, ***p < 0.001.

with-power and without-power grouping (Fig. 2). The results show that when participants didn't earn rewards through danmaku voting, those with jury-based content moderation power significantly and positively predicted social identity (simple slope = 0.02, $M_{with-power} = 5.61$, SE = 0.11; $M_{without-power} = 5.27$, SE = 0.11; p = 0.025), with a 95% confidence interval of [0.05,0.56]; when participants earned rewards through danmaku voting, the predictive effect of jury-based content moderation power on social identity remained significant and was enhanced (simple slope = 0.48, $M_{with-power} = 6.09$, SE = 0.11; $M_{without-power} = 5.29$, SE = 0.11; p = 0.000), with 95% confidence interval [0.43,0.99]. This indicates that rewards positively moderated the relationship between jury-based content moderation and social identity, supporting H3a.

3.2.2.2. Moderated mediating effects. 5000 bootstraps with a 95% confidence interval were employed to estimate the moderated mediation relationships using PROCESS Model 7 (Hayes, 2009). According to Table 3, reputation rewards moderate the indirect effect of jury-based content moderation power on user stickiness mediated by social identity, with a moderated mediation index of 0.40 and a confidence interval of [0.03, 0.78], affirming the effect's significance, supporting H3b.

3.3. Experiment 3: Moderating effect of power scarcity

The third experiment tested the moderating effect of power scarcity on the association between jury-based content moderation power and social identity and the moderated mediation effect on user stickiness via social identity (i.e., H4a and H4b), using a 2 (with/without jury-based content moderation power) × 2 (high/low power scarcity) between-groups design. The sample size calculation using G*Power software is consistent with Experiment 2.

3.3.1. Procedure

The experiment recruited 240 new subjects (not involved in the previous studies) and randomly divided them into four groups: with-power-high-scarcity, with-power-low-scarcity, without-power-high-scarcity, and without-power-low-scarcity. The identity and feedback manipulations in Experiment 1 were used to differentiate with-power and without-power groups. To manipulate perceived power scarcity, we displayed the number of invited participants for voting alongside the identity messages. Specifically, the high-scarcity groups were informed: "*Only 30 jurors/investigators have been invited for danmaku governance/survey*", while the low-scarcity groups were told: "*A total of 10,000 jurors/investigators have been invited for danmaku governance/survey*". These specific numbers for high/low power scarcity were determined after consulting platform management experts.

Finally, subjects completed scales measuring user stickiness (Cronbach's α = 0.838), social identity (Cronbach's α = 0.921), and perceived control over jury-based content moderation power. All scales used 7-point Likert scales. We also conducted manipulation checks for power scarcity (Park et al., 2022) (3 items, e.g., "I feel few users were invited as investigators/jurors for the research agency/video platform. ") and collected demographic information. Each participant received 7 RMB upon completion.

3.3.2. Results

For data validity, we excluded samples that failed attention checks (n = 17, 7.08%), had excessively brief response times (n = 1, 0.42%), or exhibited patterned responses (n = 3, 1.25%). We collected 219 valid questionnaires (132 female; majority aged 21–40; most holding an undergraduate degree), achieving a 91.25% effective response rate. A Harman's single-factor test was conducted as in Study 1, confirming no significant common method bias.

Manipulation checks via independent samples T-tests confirmed that both perceived control (M = 6.07 vs. 4.98, p < 0.01) and power scarcity (M = 5.25 vs. 2.58, p < 0.01) were successful manipulated. The results of two-way ANOVA (Table 4) reveal that participants with jury-based content moderation power have a significantly higher user stickiness compared to those without power ($M_{with-power} = 5.67$, SD = 0.94; $M_{without-power} = 5.12$, SD = 1.00, p < 0.001), consistent with Experiment 1.

3.3.2.1. Moderating effect of power scarcity. The results of two-way ANOVA (Table 4) reveal a significant impact of the interaction between jury-based content moderation power and power scarcity on users' social identity (p < 0.05), confirming the moderating effect of power scarcity on the effect of jury-based content moderation power on social identity. To examine this moderating effect in

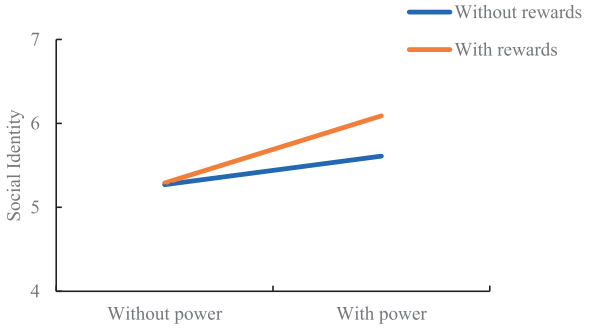


Fig. 2. Moderating effect of reputation rewards.

Table 3
Coefficients for moderated mediating effect of reputation rewards.

Jury-based Moderation Power $R^2 = 0.15$, $F = 12.48$, $p = 0.000$		Social Identity (Mediator) β	SE	t	p
Constant		5.36	0.53	10.08	0.000***
Jury-based Moderation Power		−0.12	0.34	−0.34	0.731
Rewards		−0.44	0.34	−1.29	0.197
Jury-based Moderation Power × Rewards		0.46	0.22	2.13	0.034*
Jury-based Moderation Power		User Stickiness(Dependent Variable)			
$R^2 = 0.57$, $F = 144.55$, $p = 0.000$		β	SE	t	p
Constant		0.58	0.30	1.96	0.052
Jury-based Moderation Power		0.06	0.09	0.60	0.549
Social Identity		0.88	0.06	15.82	0.000***
Indirect Effect					
	Rewards (Moderator)	Effect	SE	LLCI	ULCI
Power → Social Identity → User Stickiness	Without rewards	0.30	0.12	0.05	0.56
Power → Social Identity → User Stickiness	With rewards	0.71	0.14	0.43	0.99
Moderated Mediation Index					
		β	SE	LLCI	ULCI
Rewards		0.40	0.19	0.03	0.78

Notes:*p < 0.05, **p < 0.01, ***p < 0.001.

Table 4
Two-way ANOVA results of moderated by power scarcity.

		User Stickiness			Social Identity		
X	W	Mean	SD	p	Mean	SD	p
Without Power	Low Scarcity	5.15	1.03	0.748	5.01	1.17	0.776
	High Scarcity	5.09	0.97		4.96	0.91	
	Total	5.12	1.00		4.98	1.04	
With Power	Low Scarcity	5.42	1.12	0.007**	5.51	0.86	0.004**
	High Scarcity	5.91	0.64		6.00	0.48	
	Total	5.67	0.94		5.76	0.73	
Total	Low Scarcity	5.28	1.08	0.092	5.26	1.05	0.065
	High Scarcity	5.50	0.92		5.48	0.90	
	Without Power	5.12	1.00	0.000***	4.98	1.04	0.000***
	With Power	5.67	0.94		5.76	0.73	
X				0.000***			0.000***
W				0.092			0.065
X * W				0.033*			0.025*

Notes:*p < 0.05, **p < 0.01, ***p < 0.001.

detail, we conducted a simple slope test and drew a simple effect analysis plot (Fig. 3) based on with-power and without-power grouping. The results show that when power scarcity is low, jury-based content moderation power significantly and positively predicted social identity (simple slope = 0.50, $M_{with-power} = 5.51$, SE = 0.12; $M_{without-power} = 5.01$, SE = 0.12; p = 0.004), with a 95% confidence interval of [0.09,0.66]; when power scarcity is high, the predictive effect of jury-based content moderation power on social identity remained significant and was enhanced (simple slope = 1.04, $M_{with-power} = 6.00$, SE = 0.12; $M_{without-power} = 4.96$, SE = 0.12; p = 0.000), with 95% confidence interval [0.54,1.01]. These findings indicate that power scarcity positively moderated the relationship between jury-based content moderation power and social identity, supporting H4a.

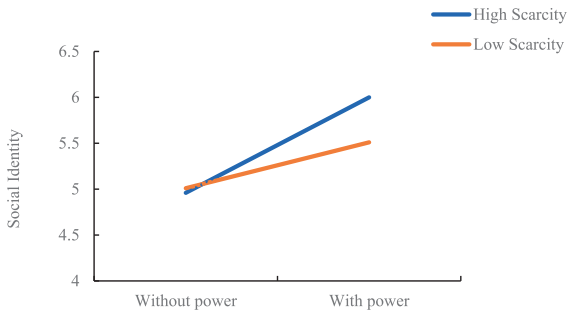


Fig. 3. Moderating effect of power scarcity.

3.3.2.2. Moderated mediating effects. We employ 5000 bootstraps with a 95% confidence interval to estimate the moderated mediation relationships using PROCESS Model 7 (Hayes, 2009). Table 5 shows that power scarcity significantly moderates the indirect effect of jury-based content moderation power on user stickiness via social identity (moderated mediation index = 0.40; confidence interval [0.05, 0.77]). The confidence interval excludes 0, confirming a significant moderated mediation effect and supporting H4b.

4. Additional analysis

While our experimental studies provide causal evidence under idealized conditions, whether it can be generalized to actual platform dynamics remains an open question. To validate the empowerment effects also manifests in a naturalistic context, we attempt to examine behavioral traces of real-world users who has participated in jury-based content moderation.

4.1. Data collection

We choose Bilibili, a leading online video platform in China, as our context for data collection. The platform launched its jury system in 2017, allowing users to moderate user-generated content. Starting in November 2017, Bilibili began to commend a few active contributors of jury-based moderation every week in reports named "Disciplinary Committee Weekly Reports". These reports offer us a unique opportunity to identify users who actively played the role of jurors in certain weeks.

We manually retrieved 128 weekly reports from October 2021 to July 2024 and obtained usernames of commended users. After dropping usernames that are not in use (possibly due to change of username), we obtained a final list of 1,343 users with accessible profile page. 391 of these users has omitted their activities, leaving us with a final sample of 952 users. For each user, we collected their activity data – including personal posts, video collections, and submitted videos – during a 24-week observational period, which covers 12 weeks before and after the week they were listed in the weekly report. Summary statistics and variable definitions are presented in Table 6.

4.2. Model specification

Based on the collected dataset, we employed a staggered difference-in-differences (DiD) design to examine the impact of participation in jury-based content moderation on user stickiness (as reflected in their platform activity). The treatment is defined as a user being commended in the weekly report for their active participation in content moderation. Observations of users who had not yet been listed in the weekly reports will act as the control groups. The event time that users started to be affected by jury-based content moderation is the week they first appeared in a report. Our empirical model is specified as follows:

$$y_{it} = \beta_0 + \beta_1 \, Commended_{it} + \theta X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

Where y_{it} represents the number of personal posts by user i in week t. $Commended_{it}$ is the key independent variable – a dummy that equals 1 if user i has already been commended for being jurors by week t, and 0 otherwise. X_{it} represents control variables (i.e., number of submitted videos). μ_i captures user fixed effects, λ_t captures week fixed effects, and ε_{it} is the error term.

4.3. Results

Table 7 presents the results and a series of robustness checks of our DiD analysis. Column (1) shows the baseline estimation results.

Table 5
Coefficients for moderated mediating effect of power scarcity.

Jury-based Moderation Power		Social Identity (Mediator)			
$R^2 = 0.19$, F = 16.89, p = 0.000		β	SE	t	p
Constant		5.10	0.60	8.45	0.000***
Jury-based Moderation Power		−0.04	0.38	−0.11	0.912
Power Scarcity		−0.59	0.38	−1.56	0.121
Jury-based Moderation Power × Power Scarcity		0.54	0.24	2.26	0.025*
Jury-based Moderation Power		User Stickiness(Dependent Variable)			
$R^2 = 0.51$, F = 111.90, p = 0.000		β	SE	t	p
Constant		1.46	0.27	5.40	0.000***
Jury-based Moderation Power		−0.01	0.10	−0.12	0.904
Social Identity		0.74	0.05	13.77	0.000***
Indirect Effects					
	Power Scarcity(Moderator)	Effect	SE	LLCI	ULCI
Power → Social Identity → User Stickiness	Low *Scarcity*	0.37	0.15	0.09	0.66
Power → Social Identity → User Stickiness	High *Scarcity*	0.76	0.12	0.54	1.01
Moderated Mediation Index					
		β	SE	LLCI	ULCI
Power Scarcity		0.40	0.19	0.05	0.77

Notes: *p < 0.05, **p < 0.01, ***p < 0.001.

Table 6
Descriptive statistics.

Variable	Description	N	Mean	SD	Min	Max
$Post_{it}$	The number of posts published by user *i* in week *t*	16,184	1.578	8.451	0	540
$Collection_{it}$	The number of videos collected by user *i* in week *t*	16,184	0.554	6.177	0	241
$Submission_{it}$	The number of videos submitted by user *i* in week *t*	16,184	0.302	1.187	0	33

Table 7
Baseline regression and robustness check.

Variables	(1)	(2)	(3)	(4)
	DV: Post Week(−8~+8)	DV: Collection Week(−8~+8)	DV: Post Week(−4~+4)	DV: Post Week(−12~+12)
Commended	0.813***	0.180***	0.476***	0.924***
	(0.103)	(0.050)	(0.094)	(0.109)
Submission	0.993***	−0.028*	0.980***	0.986***
	(0.069)	(0.014)	(0.075)	(0.064)
Constant	−0.662	0.697*	−0.728***	−0.576
	(0.986)	(0.274)	(1.868)	(0.779)
Observations	16,184	16,184	8568	23,800
User fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.182	0.064	0.182	0.171

Notes: *p < 0.05, **p < 0.01, ***p < 0.001.

The coefficient of $Commended_{it}$ is positive and significant ($\beta = 0.813$, $p < 0.001$), indicating that users create significantly more posts after the week they were reported for being active in content moderation. This provides real-world evidence on the impact of jury-based content moderation on user stickiness.

To validate the robustness of results, we conducted several checks. We first replace the dependent variable (i.e., number of posts) with another user activity − number of video collections. As shown in Column (2), the results are consistent with our baseline model, confirming that the findings hold for other measure of user stickiness. We also examined the stability of our results by varying the observation window. While the baseline model employs a 16-week window (8 weeks before and after the event), we also conduct analysis on a shorter 8-week window (−4 to + 4 weeks) and longer 24-week window (−12 to + 12 weeks). The results, as presented in Column (3) and Column (4) respectively, are consistent with our baseline model.

We also employed an event study approach to test the parallel pre-trend assumption of DiD design. Fig. 4 presents the results of our event study approach. As shown in the figure, the coefficients for the pre-treatment periods are close to zero with no clear pre-trend, while the coefficients for the after-treatment periods are significantly greater than zero. This suggests that users' platform activity remain stable prior to their active participation in content moderation and increase sharply after the event. The results validate our main findings and support that the parallel pre-trends assumption is not violated.

Since all users in our sample are eventually exposed to jury-based content moderation, we also employ Regression Discontinuity in Time (RDiT) as a robustness check to mitigate potential biases from two-way fixed effects in the staggered DiD design. The estimated jump in dynamic posts at the cutoff (i.e., the week when being listed in the weekly reports) is both positive and significant (see Appendix B for the full specification and results), remaining consistent to the finding of our DiD design.

Overall, the results of our additional analysis provide consistent evidence that active participation in jury-based moderation

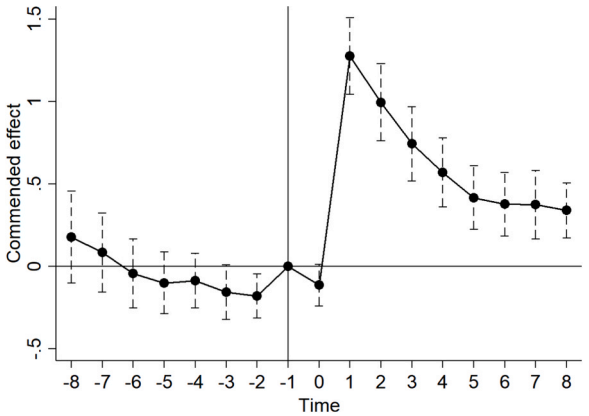


Fig. 4. Parallel trends based on event study approach.

significantly increases subsequent user engagement, and the results are robust to variations in outcome measures, sample windows, and models.

5. Conclusion and discussion

5.1. Research conclusions

Based on social identity theory, this paper reveals that jury-based content moderation positively influences user stickiness, indicating that this superior power and authority not only bring users a stronger sense of control but also increase their loyalty and dependence on the platform. Additionally, social identity serves as a mediator in this relationship. As shown in Experiment 1, users engaged in jury-based content moderation exhibit higher social identity, which in turn enhances their platform stickiness and willingness to remain active. This suggests that users' sense of identification serve as a crucial mechanism linking the jury-based content moderation and user continued engagement. Furthermore, Experiment 2 demonstrates that reputation rewards amplify the positive effect of jury-based content moderation on social identity, while Experiment 3 reveals that power scarcity strengthens the positive effect of jury-based content moderation on social identity. These findings imply that jury-based content moderation can better enhance users' identification and platform stickiness when accompanied by extrinsic and intrinsic benefits. To test the generalizability of these findings beyond the lab, we further analyzed actual platform behavior and found that users increased platform engagement after actively participated in content moderation (as recorded in the platforms weekly reports), reinforcing the empowerment effect observed in our experiments.

5.2. Theoretical implications

This study contributes to platform stickiness research by deepening our understanding of user empowerment participation models and offering new empirical insights. First, existing research on jury-based content moderation has largely examined its governance outcomes, such as improving decision legitimacy and influencing user behavior (Fan and Zhang, 2020; Drolsbach et al., 2024; Hu et al., 2021; Chuai et al., 2024), which has focused primarily on users being judged. Our study acknowledges platform users' tripartite identity as content consumers, content producers and content managers, thus shifting the attention to users carrying out the jury-based content moderation. By uncovering how users' social identity perception mediates the relationship between jury-based content moderation power and platform stickiness, we provide novel insights into the mechanisms underlying the empowerment effect of crowd-sourced governance approach. While this kind of jury-based systems is often regarded as a content moderation tool that satisfies users through preventing harmful content and increasing decision legitimacy, it could also function as a user retention strategy that facilitates users' psychological attachment to the platform through empowering them.

Second, previous research on user stickiness primarily focused on platform-driven technical measures, such as interface optimization, personalized services, and interactive functionalities (Li et al., 2024; Li et al., 2021; Pang et al., 2024). These studies emphasize on one-way technological provision, which highlights the importance of providing instrumental or hedonic value but often overlooks the potential of users' active agency. To enrich our theoretical understanding of user stickiness enhancement, our study shifts the focus from the platform-driven approach to a community-driven approach. Specifically, we uncover the critical role of social identity perception in mechanism linking user empowerment and user stickiness, which extends our understanding of user retention strategy beyond traditional technical or hedonic features.

Third, while self-determination theory has been widely applied in motivation and behavior research, its application in the context of platform governance remains limited. Our study extends the theory by identifying reputation rewards and power scarcity as boundary conditions that moderate the effect of user empowerment on stickiness. This enriches the understanding of how both external (social recognition) and internal (self-worth) motivations condition the outcomes of participatory governance, offering a nuanced account of when and for whom empowerment works best in digital environments.

Overall, by integrating social identity theory and self-determination theory, this study provides a novel theoretical perspective and empirical evidence on the interplay between user empowerment, social identity, and incentive strategies (both external and internal) and how they collectively contribute to enhanced platform stickiness.

5.3. Managerial implications

This research provides practical insights for enhancing user stickiness through social media platform governance. First, our study highlights that jury-based content moderation is not merely a governance mechanism, but also a powerful user empowerment tool that can strengthen users' psychological attachment to the platform. This suggests that platforms should not only view such systems as instruments to improve content quality, but also as means to engage users as co-governors. Actively involving users in moderation empowers them with a sense of agency, reinforcing their identification with the platform and encouraging longer-term engagement. Managers should therefore design these systems to make users feel that their decisions have real consequences, and that they are trusted participants in maintaining community norms.

Second, a key finding of our study is the important role of user empowerment (i.e., jury-based content moderation power) in fostering user stickiness via creating stronger self-identification with the platform. This suggests that platforms should involve users in content governance activities, which can enhance users' sense of ownership and responsibility. For instance, platforms may consider implementing jury-based content moderation mechanisms like “Discipline Commission”, “Community Courts” or “Crowd Jury”

systems, as seen in many online platforms like Bilibili and Weibo, to incorporate users into content review and evaluation. Notably, a crucial aspect of such empowerment is making users feel valued and influential in the governance process. Hence, platform should craft the mechanism to ensure that jury-based content moderation participants are aware of their decisions' impact.

Third, the study highlights the importance of external incentivization (e.g., reputation rewards) in boosting the effectiveness of jury-based content moderation. This indicates that platforms utilizing jury-based content moderation should also consider implementing diverse reward schemas, including honors, badges, superior status, and social recognition, to strengthen users' sense of identity and belonging. For example, platforms can award “Contributor Badges” or create a “Reputation System”, as seen in many online UGC platforms like GitHub or Stack Overflow, to recognize users' efforts and contributions. This could encourage and sustain users' engagement in the platform.

Finally, our findings also suggest that strategic management of jury-based content moderation power through controlled scarcity can enhance its effectiveness. Platforms should implement methods that can create a sense of exclusivity-such as setting participant qualification thresholds, adopting an invitation-based system, or restricting the number of judges involved in each evaluation task-to enhance the perceived value and meaningfulness of jury-based content moderation power. It is essential that all efforts related to power scarcity should be visible and perceivable to users. These strategies can motivate active participation of users without significantly increasing platform costs.

Collectively, these practical recommendations demonstrate how platform managers could leverage jury-based content moderation mechanisms, incentivization, and exclusivity strategies to foster long-term user retention and platform sustainability.

6. Limitations

Despite achieving meaningful insights into the relationship between user empowerment in platform governance and user stickiness, this study has some limitations. Firstly, while we utilize the Disciplinary Committee Weekly Reports from Bilibili to identify users who have actively engaged in jury-based content moderation, this approach has inherent limitations. Specifically, the reports only record a subset of exceptionally active jurors. This may constrain our identification to users with intensive participation, thus limiting the generalizability of the findings to the broader user base. Furthermore, we cannot fully ascertain whether these users had previously engaged in moderation before being reported, making it difficult to establish a clean pre-moderation baseline. The public commendation itself may also act as a reputational incentive, potentially confounding the pure empowerment effect with visibility- or reward-based motivations. However, we believe these limitations can be mitigated by the consistent findings revealed in our controlled experiments. Second, despite the effort made to simulate real scenarios, our online experiments cannot perfectly mimic the users' experience in real-world settings and may face external validity concerns. Future research could validate these findings using field experiments and case studies. Third, the use of danmaku video websites as our research context may limit generalizability to other social media platforms, suggesting the need to explore diverse platform types. Lastly, this study examines only two moderating variables—reputation rewards and power scarcity. Future studies could investigate additional factors, such as user traits and platform cultural characteristics, to deepen understanding of user participation mechanisms.

In summary, this study provides new insights into the relationship between user participation in platform governance and stickiness, highlighting the mechanisms that drive this connection. It emphasizes the unique role of social media users and underscores the importance of user agency in shaping platform stickiness. By demonstrating how user active involvement in governance enhances user experience and platform loyalty, the study advances a community-driven perspective of platform governance. Future research can build on these findings to deepen understanding and improve governance strategies that prioritize user experience and engagement.

CRediT authorship contribution statement

Baojun Ma: Supervision, Methodology, Funding acquisition, Conceptualization. **Xiaoyan Wen:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Yi Chen:** Writing – review & editing, Supervision, Conceptualization. **Yao Mu:** Supervision, Conceptualization. **Yi Huang:** Methodology.

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.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China [Grants 72172092, 72302153, 72472103, 72202140], Shanghai Key Laboratory of Brain-Machine Intelligence for Information Behavior [22dz2261100], Fundamental Research Funds for the Central Universities [Grant 41005067], Humanities and Social Science Fund of Ministry of Education of China [Grant 22YJCZH064].

Appendix A:. Details of items in measurement scales and original sources

.

Table 1
Measurement Scales, Items, and Sources.

Scale	Measurement Items	Source
Stickiness	– I intend to continue using this online video platform. – I intend to continue using this online video platform instead of other alternative services or platforms. – I expect to continue using this online video platform in the future. –I will increase the frequency of using this online video platform.	Huang & Chung (2024); Li et al. (2024)
Social Identity	– I see myself as a representative member of this online video platform. –I see myself as a member of this online video platform. – I identify myself as a member of this online video platform. – I feel a sense of belonging to this online video platform. – I am proud to be a member of this online video platform. –I enjoy being a member of this online video platform. – I have a strong emotional connection with this online video platform. –I am a valuable member of this online video platform. –I am an important member of this online video platform. –I am a high-quality member of this online video platform.	Bagozzi & Dholakia (2006); Farivar & Wang (2022)
Perceived Control	– During the danmaku display, my vote would be accepted by the video platform. – My vote is important for the platform's governance. – My vote would be reflected in the platform's final decisions. –My vote would influence the platform's governance strategy.	Chun & Lee (2017); Huh et al. (2023)
Reputation Rewards	– After participating in the danmaku quality survey/governance, I received an honorary reward from the research agency/video platform.	McKernan et al. (2015)
Power Scarcity	– I feel few investigators/jurors in the danmaku survey/governance. – I feel the number of investigators/jurors in the survey/governance was scarce. – I feel few users were invited as investigators/jurors for the research agency/video platform.	Park et al. (2022)

Appendix B:. Additional analysis using regression discontinuity in time (RDiT)

We conduct an additional robustness check using a Regression Discontinuity in Time (RDiT) approach. While the DiD strategy leverages untreated users in the same week as controls, the RDiT model focuses exclusively on within-user variation before and after the commendation week, treating the week of commendation as a sharp temporal cutoff.

This specification allows us to test whether user behavior exhibits a local structural change around the shock, independent of the control group comparison. We adopt a parametric RDiT model with user fixed effects and polynomial time trends as specified below: $y_{it} = \beta_0 + \beta_1 Commended_t + \beta_2 Duration_t + \beta_3 Commended_t \times Duration_t + \theta X_{it} + \mu_i + \varepsilon_{it}$ where $y_{it}$is the number of dynamic posts by user i in week t. $Commended_t$ is a dummy variable that equals 1 (0) if week t is after (before) the commendation. $Duration_t$ is the number of weeks after or before the commendation. The interaction term $Commended_t \times Duration_t$ allows the regression function to differ on either side of the cutoff point. X_{it} is time-varying controls (specifically, submissions). $\mu_i$captures user fixed effects, and $\varepsilon_{it}$is the error term.

We first conducted a paired t-test comparing user behavior before and after commendation. Results show a significant increase in dynamic posting activity ($M_{pre} = 1.404$, $SE = 0.085$; $M_{post} = 2.793$, $SE = 0.177$, $p = 0.000$).

Subsequently, we estimate the RDiT model. As shown in Table 2, users significantly increased their dynamic posting after being commended. These findings further confirm the incentivizing effect of commendation on user engagement behaviors on the platform.

Table 2
Result of RDiT Model.

Variables	DV: Post
Commended	1.304***
	(0.146)
Duration	−0.077**
	(0.023)
Commended × duration	−0.058
	(0.213)
Submission	0.952***
	(0.078)
Constant	1.316***
	(0.075)
Observations	7616
User fixed effects	Yes
Adjusted R^2	0.171

Notes: *p < 0.05, **p < 0.01, ***p < 0.001.

Data availability

Data will be made available on request.

References

Acar, O.A., Puntoni, S., 2016. Customer empowerment in the digital age. J. Advert. Res. 56 (1), 4–8. https://doi.org/10.2501/JAR-2016-007.

Bagozzi, R.P., Dholakia, U.M., 2006. Antecedents and purchase consequences of customer participation in small group brand communities. Int. J. Res. Mark. 23 (1), 45–61. https://doi.org/10.1016/j.ijresmar.2006.01.005.

Chen, M.-H., Tsai, K.-M., 2020. An empirical study of brand fan page engagement behaviors. Sustainability 12 (1), 434. https://doi.org/10.3390/su12010434.

Chuai, Y., Tian, H., Pröllochs, N., et al., 2024. Did the roll-out of Community Notes reduce engagement with misinformation on X/Twitter? Proc. ACM Hum.-Comput. Interact. 8 (CSCW2). https://doi.org/10.1145/3686967. Article 428.

Chun, J.W., Lee, M.J., 2017. When does individuals' willingness to speak out increase on social media? Perceived social support and perceived power/control. Comput. Hum. Behav. 74, 120–129. https://doi.org/10.1016/j.chb.2017.04.010.

Deci, E.L., Olafsen, A.H., Ryan, R.M., 2017. Self-determination theory in work organizations: the state of a science. Annu. Rev. Organ. Psychol. Organ. Behav. 4 (1), 19–43. https://doi.org/10.1146/annurev-orgpsych-032516-113108.

Deci, E.L., Ryan, R.M., 2013. Intrinsic motivation and self-determination in human behavior. Springer.

Dholakia, U.M., Bagozzi, R.P., Pearo, L.K., 2004. A social influence model of consumer participation in network- and small-group-based virtual communities. Int. J. Res. Mark. 21 (3), 241–263. https://doi.org/10.1016/j.ijresmar.2003.12.004.

Drolsbach, C.P., Solovev, K., Pröllochs, N., 2024. Community notes increase trust in fact-checking on social media. PNAS Nexus 3 (7). https://doi.org/10.1093/pnasnexus/pgae217.

Ejaz, N., Razi, F., Choudhury, S., 2024. Towards comprehensive cyberbullying detection: a dataset incorporating aggressive texts, repetition, peerness, and intent to harm. Comput. Hum. Behav. 153, 108123. https://doi.org/10.1016/j.chb.2023.108123.

Elsharnouby, M.H., Mohsen, J., Saeed, O.T., et al., 2021. Enhancing resilience to negative information in consumer–brand interaction: the mediating role of brand knowledge and involvement. J. Res. Interact. Mark. 15 (4), 571–591. https://doi.org/10.1108/JRIM-05-2020-0107.

Farivar, S., Wang, F., 2022. Effective influencer marketing: a social identity perspective. J. Retail. Consum. Serv. 67, 103026. https://doi.org/10.1016/j.jretconser.2022.103026.

Faul, F., Erdfelder, E., Lang, A.-G., et al., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191. https://doi.org/10.3758/BF03193146.

Gao, X., Yee, C.-L., Choo, W.-C., 2022. How attachment and community identification affect user stickiness in social commerce: a consumer engagement experience perspective. Sustainability 14 (20), 13633. https://doi.org/10.3390/su142013633.

Gorwa, R., Binns, R., Katzenbach, C., 2020. Algorithmic content moderation: technical and political challenges in the automation of platform governance. Big Data Soc. 7 (1). https://doi.org/10.1177/2053951719897945.

Grant, A.M., 2008. The significance of task significance: job performance effects, relational mechanisms, and boundary conditions. J. Appl. Psychol. 93 (1), 108–124. https://doi.org/10.1037/0021-9010.93.1.108.

Gu, T., Cheng, Z., Zhang, Z., et al., 2022. Formation mechanism of contributors' self-identity based on social identity in online knowledge communities. Front. Psychol. 13, 1046525. https://doi.org/10.3389/fpsyg.2022.1046525.

Hayes, A.F., 2009. Beyond Baron and Kenny: statistical mediation analysis in the new millennium. Commun. Monogr. 76 (4), 408–420. https://doi.org/10.1080/03637750903310360.

Fan, J., Zhang, A.X., 2020. Digital juries: A civics-oriented approach to platform governance. Proc. CHI Conf. Hum. Factors Comput. Syst. https://doi.org/10.1145/3313831.3376293.

Hettiachchi, D., Goncalves, J., 2020. Towards effective crowd-powered online content moderation. Proc. Australas. Comput.-Hum. Interact. Conf. https://doi.org/10.1145/3369457.3369491.

Hsu, C.-L., Chen, Y.-C., Yang, T.-N., et al., 2017. Do website features matter in an online gamification context? focusing on the mediating roles of user experience and attitude. Telemat. Inform. 34 (4), 196–205. https://doi.org/10.1016/j.tele.2017.01.009.

Hu, X.E., Whiting, M.E., Bernstein, M.S., 2021. Can online juries make consistent, repeatable decisions? Proc. CHI Conf. Hum. Factors Comput. Syst. https://doi.org/10.1145/3411764.3445433.

Huang, T.-L., Chung, H.F., 2024. Impact of delightful somatosensory augmented reality experience on online consumer stickiness intention. J. Res. Interact. Mark. 18 (1), 6–30. https://doi.org/10.1108/JRIM-07-2022-0213.

Huh, J., Kim, H.-Y., Lee, G., 2023. “Oh, happy day!” Examining the role of AI-powered voice assistants as a positive technology in the formation of brand loyalty. J. Res. Interact. Mark. 17 (5), 794–812. https://doi.org/10.1108/JRIM-10-2022-0328.

Lampe, C., Zube, P., Lee, J., et al., 2014. Crowdsourcing civility: a natural experiment examining the effects of distributed moderation in online forums. Gov. Inf. Q. 31 (2), 317–326. https://doi.org/10.1016/j.giq.2013.11.005.

Li, G., Zhao, Z., Li, L., et al., 2024. The relationship between AI stimuli and customer stickiness, and the roles of social presence and customer traits. J. Res. Interact. Mark. 18 (1), 38–53. https://doi.org/10.1108/JRIM-07-2022-0222.

Li, S., Zhu, B., Zhu, H., et al., 2024. Consumer sticky purchase intention for personalized digital collections of cultural heritage in metaverse: from the perspective of self-concept. Electron. Commer. Res. (advance Online Publication) 1–39. https://doi.org/10.1007/s10660-024-09927-1.

Li, Y., Li, X., Cai, J., 2021. How attachment affects user stickiness on live streaming platforms: a socio-technical approach perspective. J. Retail. Consum. Serv. 60, 102478. https://doi.org/10.1016/j.jretconser.2021.102478.

Lin, S.-W., Huang, C.-D., 2024. Hooked on audio! Unveiling the secrets of podcast stickiness through social identity and uses and gratification theories. Technol. Soc. 76, 102422. https://doi.org/10.1016/j.techsoc.2023.102422.

McKernan, B., Martey, R.M., Stromer-Galley, J., et al., 2015. We don't need no stinkin' badges: the impact of reward features and feeling rewarded in educational games. Comput. Hum. Behav. 45, 299–306. https://doi.org/10.1016/j.chb.2014.12.028.

Micek, C., Solovey, E.T., 2024. Examining the impact of digital jury moderation on the polarization of U.S. political communities on social media. Interact. Comput. 37 (4), 234–252. https://doi.org/10.1093/iwc/iwae036.

Mostafa, R.B., Sobhy Temerak, M., 2024. Does consumer empowerment enhance brand page stickiness? the role of brand page experience and brand love. J. Res. Interact. Mark. 18 (6), 1136–1154. https://doi.org/10.1108/JRIM-06-2023-0192.

Mou, Y., Ma, Y., Guo, D., et al., 2025. Effect of platform gamification rewards on user stickiness. Manag. Decis. 63 (3), 824–849. https://doi.org/10.1108/MD-09-2023-1688.

Pang, H., Ruan, Y., Zhang, K., 2024. Deciphering technological contributions of visibility and interactivity to website atmospheric and customer stickiness in AI-driven websites: the pivotal function of online flow state. J. Retail. Consum. Serv. 78, 103795. https://doi.org/10.1016/j.jretconser.2024.103795.

Pang, H., Zhang, K., 2024. How multidimensional benefits determine cumulative satisfaction and eWOM engagement on mobile social media: reconciling motivation and expectation disconfirmation perspectives. Telemat. Inform. 93, 102174. https://doi.org/10.1016/j.tele.2024.102174.

Park, J., Eom, H.J., Spence, C., 2022. The effect of perceived scarcity on strengthening the attitude–behavior relation for sustainable luxury products. J. Prod. Brand Manag. 31 (3), 469–483. https://doi.org/10.1108/JPBM-09-2020-3091.

Podsakoff, P.M., Mackenzie, S.B., Lee, J.-Y., et al., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88 (5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879.

Preece, J., Shneiderman, B., 2009. The reader-to-leader framework: motivating technology-mediated social participation. AIS Trans. Hum.-Comput. Interact. 1 (1), 13–32. https://aisel.aisnet.org/thci/vol1/iss1/5.
Ren, Y., Kraut, R., Kiesler, S., 2007. Applying common identity and bond theory to design of online communities. Organ. Stud. 28 (3), 377–408. https://doi.org/10.1177/0170840607076007.
Rong, K., Xiao, F., Zhang, X., et al., 2019. Platform strategies and user stickiness in the online video industry. Technol. Forecast. Soc. Change 143, 249–259. https://doi.org/10.1016/j.techfore.2019.01.023.
Seering, J., 2020. Reconsidering self-moderation: the role of research in supporting community-based models for online content moderation. Proc. ACM Hum.-Comput. Interact. 4 (CSCW2). https://doi.org/10.1145/3415178. Article 107.
Shams, R., Chatterjee, S., Chaudhuri, R., 2024. Developing brand identity and sales strategy in the digital era: moderating role of consumer belief in brand. J. Bus. Res. 179, 114689. https://doi.org/10.1016/j.jbusres.2024.114689.
Shi, C., Hu, P., Fan, W., et al., 2024. Competitive peer influence on knowledge contribution behaviors in online Q&A communities: a social comparison perspective. Internet Res. 34 (5), 1577–1601. https://doi.org/10.1108/INTR-07-2022-0510.
Singh, S., Singh, N., Kalinić, Z., et al., 2021. Assessing determinants influencing continued use of live streaming services: an extended perceived value theory of streaming addiction. Expert Syst. Appl. 168, 114241. https://doi.org/10.1016/j.eswa.2020.114241.
Spence, R., Bifulco, A., Bradbury, P., et al., 2023. The psychological impacts of content moderation on content moderators: a qualitative study. Cyberpsychol 17 (4). https://doi.org/10.5817/CP2023-4-8.
Sun, Y., Fang, Y., Lim, K.H., 2012. Understanding sustained participation in transactional virtual communities. Decis. Support Syst. 53 (1), 12–22. https://doi.org/10.1016/j.dss.2011.10.006.
Tajfel, H., Turner, J., 1979. An integrative theory of intergroup conflict. Soc. Psychol. Intergroup Relations 33, 94–109.
Taylor, S.E., Lobel, M., 1989. Social comparison activity under threat: downward evaluation and upward contacts. Psychol. Rev. 96 (4), 569–575. https://doi.org/10.1037/0033-295X.96.4.569.
Wachs, S., Wright, M.F., Gámez-Guadix, M., 2024. From hate speech to HateLess: the effectiveness of a prevention program on adolescents' online hate speech involvement. Comput. Hum. Behav. 157, 108250. https://doi.org/10.1016/j.chb.2024.108250.
Wang, J., 2021. How and why people are impolite in danmu? Internet Pragmat. 4 (2), 295–322. https://doi.org/10.1075/ip.00057.wan.
Wang, S., 2023. Factors related to user perceptions of artificial intelligence (AI)-based content moderation on social media. Comput. Hum. Behav. 149, 107971. https://doi.org/10.1016/j.chb.2023.107971.
Wasko, M.M., Faraj, S., 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Q. 29 (1), 35–57. https://doi.org/10.2307/25148667.
Wei, K., 2023. The impact of barrage system fluctuation on user interaction in digital video platforms: a perspective from signaling theory and social impact theory. J. Res. Interact. Mark. 17 (4), 602–619. https://doi.org/10.1108/JRIM-06-2022-0160.
Weinstein, M., 2022. The power of scarcity: Leveraging urgency and demand to influence customer decisions. McGraw Hill.
Werner, D., Adam, M., Benlian, A., 2022. Empowering users to control ads and its effects on website stickiness. Electron. Mark. 32 (3), 1373–1397. https://doi.org/10.1007/s12525-022-00576-6.
Xu, J., Ramayah, T., Arshad, M.Z., et al., 2025. Decoding digital dependency: flow experience and social belonging in short video addiction among middle-aged and elderly chinese users. Telemat. Inform. 96, 102222. https://doi.org/10.1016/j.tele.2024.102222.
Zhang, F., Zhang, H., Gupta, S., 2023. Investor participation in reward-based crowdfunding: impacts of entrepreneur efforts, platform characteristics, and perceived value. Inf. Technol. Manag. 24 (1), 19–36. https://doi.org/10.1007/s10799-022-00363-x.
Zhang, J., Liu, R., 2024. Why do chinese people consume video game live streaming on the platform? an exploratory study connecting affordance-based gratifications, user identification, and user engagement. Telemat. Inform. 86, 102075. https://doi.org/10.1016/j.tele.2023.102075.
Zhao, A., Hobbs, W., 2025. The effects and non-effects of social sanctions from user jury-based content moderation decisions on Weibo. Proc. CHI Conf. Hum. Factors Comput. Syst. https://doi.org/10.1145/3706598.3713154.
Zhao, A., Hu, H., 2025. Unveiling strategic governance and user dynamics in Weibo's community-driven content moderation system. J. Quant. Descr. Digit. Media 5. https://doi.org/10.51685/jqd.2025.013.