

RESEARCH



# “People are Way too Obsessed with Rank”: Trust System in Social Virtual Reality

Qijia Chen<sup>\*1</sup>, Jie Cai<sup>2</sup> & Giulio Jacucci<sup>1</sup>

<sup>\*1</sup>*Department of Computer Science, University of Helsinki, Helsinki, Finland (E-mail: qijia.chen@helsinki.fi);* <sup>2</sup>*College of Information Sciences and Technology, Pennsylvania State University, Pennsylvania, USA*

Accepted: 27 March 2024

**Abstract.** Social Virtual Reality (VR) is growing in popularity and has drawn the attention of HCI academics. Social VR experiences harassment just like other online environments. The Trust System (TS) in VRChat, one of the most prominent social VR platforms, is designed to measure and indicate users’ trustworthiness in order to reduce toxicity in the platform. In this research, we analyzed data from “r/VRChat,” to understand how users perceive the system. We found that users interpret the system differently. Problems in its implementation cause distrust. The trust ranks, while intended to promote positive interactions, can actually lead to stereotyping and discourage communication between users of different ranks. The hierarchical structure within the ranks exacerbates discrimination and conflicts, particularly against the low-ranked users. We further discuss that trust ranks present challenges to newcomers and contribute to a competitive atmosphere that hinders the formation of less toxic norms. Finally, we provide implications for the future design of similar systems.

**Keywords:** Social virtual reality, Reputation system, Toxicity, Content moderation

## 1 Introduction

Social Virtual Reality (VR) is steadily gaining popularity and has become a focal area of interest for Human–Computer Interaction (HCI) scholars. Social VR refers to three-dimensional, immersive environments where individuals interact and socialize using head-mounted devices (McVeigh-Schultz et al. 2018). Users are represented by avatars, controlled through body tracking technologies (Freeman et al. 2022a), facilitating lifelike verbal and non-verbal interactions (Li et al. 2019; Wang et al. 2019).

However, the rise of social VR has brought with it a challenge: harassment (Blackwell et al. 2019; Shriram and Schwartz 2017; Rachel 2022; Freeman et al. 2022b; Frenkel and Browning 2021). Studies indicate that harassment in social VR is more severe compared to other social media platforms (Blackwell et al. 2019; Rachel 2022). This severity is attributed to VR’s distinctive qualities, such

as immersive multi-modal communication (encompassing voice, gesture, proxemics, gaze, and facial expression) (Blackwell et al. 2019). The sense of embodiment and presence in VR intensifies the harassment experience, making it more acute than in other computer-mediated environments (Slater et al. 2009). VRChat, a leading social VR platform, has notably struggled with harassment issues. A study by the non-profit Center for Countering Digital Hate, which involved monitoring VRChat user activity for over 11 h, found numerous instances of behavior that contravened Meta’s VR standards, including sexual harassment and abuse<sup>1</sup>.

In response, social VR platforms have implemented various measures to curb harmful behaviors. These include creating personal space bubbles, muting or blocking disruptive users, and employing automated moderation systems (Zheng et al. 2023). VRChat, for instance, has introduced a reputation system called the Trust System (TS)<sup>2</sup>. This system assesses and displays user trustworthiness through trust ranks, aiming to reduce toxicity on the platform. The TS is a rather unique approach considering other social VR platforms, and it is currently not clear *how users perceive the Trust System*.

Understanding how users perceive this system is crucial for refining safety tools and moderation practices in social VR. As the use of social VR grows, the need to evolve these tools becomes more apparent. As Bill Stillwell, product manager for VR integrity at Meta, said, “We will continue to make improvements as we learn more about how people interact in these spaces.”<sup>3</sup>

In addition, reputation systems have been used in online platforms to manage toxic behavior (League of Legends Wiki. n.d; Vidal 2023) by assigning points or ranks that encourage positive interactions and discourage negative ones (Hendrikx et al. 2015). However, research on how such systems are perceived in reducing toxicity, particularly in those platforms designed for social interaction, remains limited. Conducting research on that may guide developers in designing more effective reputation systems that encourage positive behavior while minimizing toxicity.

With that in mind, we gathered discussion data from the “r/VRChat”,<sup>4</sup> one of the largest VRChat-related online communities. Through inductive thematic analysis, we found that the logic of mitigating toxicity shown by this system is perceived as reasonable by users. However, problems in its implementation, such as the opacity of the algorithm and the inconsistency of the output, cause distrust. Furthermore, while TS is created to help combat harassment, it may inadvertently generate more toxicity. Specifically, trust ranks visibly categorize people into groups, increase the grounds people use to form stereotypes and discourage

<sup>1</sup> <https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/>

<sup>2</sup> <https://docs.VRChat.com/docs/VRChat-safety-and-trust-system>

<sup>3</sup> <https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/>

<sup>4</sup> <https://www.reddit.com/r/VRchat/>

“People are Way too Obsessed with Rank”: Trust System in Social...

communication among users in different ranks. The power hierarchy formed around the ranks can fuel discrimination and conflicts, particularly against those with a low rank (e.g., newcomers who are important for the community’s growth). Ultimately, trust ranks enhance the competitiveness among users and strengthen users’ impression that social VR is a form of a game, thus hindering the formation of less toxic norms.

## 2 Background

### 2.1 Social VR and toxicity

Over the past few years, social VR has been gaining traction quickly. Social VR commonly refers to 3 dimensional, immersive digital environments where people can interact, communicate, and socialize with each other through head-mounted devices (McVeigh-Schultz et al. 2018). Popular VR platforms at the time of writing include VRChat<sup>5</sup>, Horizon Worlds<sup>6</sup>, RecRoom<sup>7</sup> and others. In these virtual worlds, users are typically virtually represented by avatars. They embody and control those avatars via body tracking technologies (Freeman et al. 2022a). Social VR provides life-like interaction. In particular, VR-mediated social interaction affords verbal and non-verbal interaction in real-time (Li et al. 2019; Wang et al. 2019; De Simone et al. 2019) and supports communication features like animated or facial tracking-based facial expressions (Kolesnichenko et al. 2019). In addition, a strong sense of body ownership induced by technological characteristics over virtual user representations makes social VR seem to extend the social and experiential qualities of traditional shared virtual spaces (Kilteni et al. 2012; Maselli and Slater 2013). Previous studies on social VR illustrate people engaging in different social activities afforded by those platforms (Barreda-Ángeles and Hartmann 2022; Sykownik et al. 2021; Maloney and Freeman 2020), to satisfy diverse social needs, such as socializing with friends, meeting strangers worldwide, creating and exploring different worlds, and gathering for social events (Barreda-Ángeles and Hartmann 2022; Sykownik et al. 2021; Maloney and Freeman 2020). Previous studies have also focused on studying specific aspects of user behaviors and experiences in social VR. Piitulainen et al. (2022) investigated social dancing in VRChat and found that social factors are the main reasons people enjoy dancing in social VR. Sykownik et al. (2022) studied selfdisclosure and found that disclosure on social VR is similar to offline communication. The relationship with others moderates self-disclosure but is also impacted by the contextual factors that social VR applications afford. Furthermore, research has tried to understand the design

---

<sup>5</sup> <https://hello.vrchat.com/>

<sup>6</sup> <https://www.oculus.com/horizon-worlds/>

<sup>7</sup> <https://recroom.com/>

strategies of social VR platforms (McVeigh-Schultz et al. 2018; Jonas et al. 2019). For example, McVeigh-Schultz et al. (2019) conducted interviews with industry experts to understand the design choices of these social VR platforms to promote prosocial behaviors and introduced a preliminary design framework that helps to form prosocial interactions in social VR. Few previous studies in social VR cover TS (Saffo et al. 2020; Zheng et al. 2023), but little focus on the system itself rather than mentioning it to give a wider picture of the VRChat platform.

Harassment is a prevalent problem in social VR (Blackwell et al. 2019; Shriram and Schwartz 2017; Freeman et al. 2022b). Research shows that two out of seven women and 21 out of 99 men reported experiencing harassment, and 42% of the users said they witnessed someone else being harassed (Shriram and Schwartz 2017). The non-profit Center for Countering Digital Hate's research shows how frequently misconduct appears on VRChat. The researchers discovered 100 potential violations of Meta's VR guidelines, including sexual harassment and abuse, after watching user behavior for 11 h and 30 min.<sup>8</sup> Social VR applications afford a high potential for toxic behavior. It has been noted that the distinctive qualities of VR technology and the affordances of multi-modal communications, including both verbal and nonverbal interactions, such as voice, gesture, proxemics, gaze, and facial expression, not only support new forms of immersive experience but also potentially increase the risk of online harassment beyond text or voice-based harassment (Blackwell et al. 2019). The experience of being harassed in VR is intensified (Slater et al. 2009) and more acute than in other computer-mediated social spaces due to embodiment and presence. Using synchronous audio in virtual reality, unlike other social media platforms, where insults are typically provided asynchronously and primarily via text, makes harassment more severe. The synchronosness of VR also causes social VR applications to mainly facilitate interactions between strangers (Blackwell et al. 2019), causing more conflicts. Additionally, toxic game culture impacts VR (Freeman et al. 2022a), and gaming is a major significant case for virtual reality technology (Blackwell et al. 2019). Furthermore, due to social VR's emphasis on supporting open virtual worlds, simulating social contexts (such as multi-user events), and drawing in a diverse user base, harassment may be felt in social VR as being even more immersive and damaging than in traditional VR, which focuses primarily on single-player games or applications (Freeman et al. 2022b). The subjective and individualized nature of users' conceptions of harassment in social VR makes content moderation in social VR challenging and reporting toxic behaviors or avoiding them difficult (Blackwell et al. 2019).

In sum, social VR can afford immersive and natural social interaction similar to the real world. Social VR platforms have a prevalence of toxicity. The multiple reasons for that include the affordances of VR technologies, the demographic of

---

<sup>8</sup> <https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/>

“People are Way too Obsessed with Rank”: Trust System in Social...

users, and the impact of gaming culture. Furthermore, there is little research on how algorithm-based TS is appropriated.

## 2.2 Content moderation and punishment

Online harassment can have severe and adverse effects on the well-being of those targeted (Uttarapong et al. 2021; Schulenberg et al. 2023) and those dealing with it (Dosono and Semaan 2019; Steiger et al. 2021). At the moment, online platforms combat harm mainly through content moderation. Content moderation is the “governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse” (Grimmelmann 2015). According to this definition, content moderation has two focal points. A thread of research focuses on how to prevent abuse with punitive moderation strategies deployed by humans, algorithms, or both, such as reporting and flagging (Kou and Gui 2021), account suspension (Chandrasekharan et al. 2017; Cai et al. 2021), and content removal (Srinivasan et al. 2019). However, recently, scholars have also pointed out the potential drawbacks of this approach, as it could lead to user discontent and dropout from the platform with disappointment and frustration (Kou 2021) and question the legitimacy of the moderation process (Pan et al. 2022). To supplement the punitive sanction mechanism, researchers suggest that the moderation system should also consider the restorative approach to managing communities and transferring violators to positive members (Schoenebeck et al. 2021; Xiao et al. 2023), such as explaining the moderation decision to improve perceived justice and fairness (Jhaver et al. 2019), educating and communicating with violators to improve users’ trust in the moderation system (Cai et al. 2021), training the mental model to understand the intentions and personalities of violators (Cai and Wohn 2021), and improving algorithm transparency (Ma and Kou 2023). However, the approach also has shortcomings, including being very labor intensive and requiring training for moderators (Xiao et al. 2022).

Another research thread focuses on facilitating cooperation, which means that the moderation system should include components to organize how community members positively engage with each other (Kim et al. 2021; Seering et al. 2019). This approach would promote prosocial behaviors by setting a good example for the rest of the community members to mimic normative behaviors (Cai et al. 2021; Seering et al. 2017). Other mechanisms include the reward system (e.g., Karma point on Reddit) and the badge system to show the support and loyalty of community members (e.g., Twitch badge system) (Cai and Wohn 2023).

In line with research about facilitating cooperation, we focus on the TS in social VR, a reputation system aimed at promoting prosocial behaviors as a strategy to foster communities and mitigate antisocial behaviors. Appropriate behavior will typically be encouraged over the long run by the incentive of a positive reputation and the disincentive of a bad one (Hendrikx et al. 2015). Reputation systems facilitate the collection, aggregation, and distribution of data about

entities, which in turn can be used to characterize and predict the future behavior of entities (Despotovic and Aberer 2004; Resnick and Zeckhauser n.d; Ruohomaa et al. 2007); they tend to assign users points or ranks, working as social corrective and reinforcing prosocial behaviors. The reputation system has been used to combat toxicity in online spaces, such as games (League of Legends Wiki. n.d; Vidal 2023). One example is the Honor System in League of Legends (League of Legends Wiki. n.d), which allows players to ‘honor’ others on their team for deserving behaviors (e.g., good teamwork). Honorable players advance to honor levels, while those who misbehave are demoted. However, new platforms bring new affordances and challenges to the moderation system (Cai et al. 2023; Jiang et al. 2019). VRChat is a special 3D platform designed for social purposes and different from other online spaces with respect to toxic and harassing behaviors (Sabri et al. 2023), such as violating of personal space considered a form of harassment (Blackwell et al. 2019). The TS, displaying individual users’ trust ranks in a highly immersive and lifelike environment to mitigate toxic behaviors, requires a deep understanding of how such design might affect users’ behaviors and moderation practices.

### 2.3 Trust system in VRChat

The common mechanisms in social VR platforms for users to protect themselves from toxic behaviors include Safety Bubbles, Reporting, Blocking, Muting, Vote-kicking, and Escaping (Zheng et al. 2023). The TS is only used in VRChat, as shown in Table 1. Trust is often used interchangeably with the word “reputation” to describe a system that evaluates and shows the reliability of its users (Jøsang et al. 2007). This type of system is also often called the “trust and reputation system” (Jøsang 2006). Trust is a directional relationship that involves the trustor and trustee. One must assume the trustor is a thinking entity that can evaluate and make decisions based on past experiences and received information. The trustee can be a person, organization, physical entity, or even a concept (in the TS in social VR, the trustee is a user). Essentially, trust is the subjective probability that an entity (A) expects another entity (B) to perform an action on which its welfare depends. Trust can be considered the trustor’s evaluation of the trustee’s

**Table 1.** Safety Tools for user to contain toxic behaviors in four social VR platforms

Name	Safety Bubble	TS	Report	Mute	Block	Votekick	Escape
AltspaceVR	✓		✓	✓	✓		✓
VRChat		✓	✓	✓	✓	✓	✓
Horizon World	✓		✓	✓	✓	✓	✓
RecRoom	✓		✓	✓	✓	✓	

## “People are Way too Obsessed with Rank”: Trust System in Social...

reliability (e.g., expressed as probability) in the context of dependence on the trustee (Jøsang 2006).

TS on VRChat is a system that evaluates the trustworthiness (i.e., reliability) of users based on their behaviors, then ranks them accordingly<sup>9</sup>. Normal ranks include Visitor, New User, User, Known User, and Trusted User, referring to Fig. 1. Newcomers start at the Visitor level and can progress to higher ranks by actively engaging in the VRChat. The progression through these ranks is determined by an algorithm. The VRChat official does not explain how the algorithm works, except to state that the trust ranks are linked with how much time a user has spent on VRChat, how much content they have contributed, friends they have made, and other factors, as well as things that do not raise trust rank: standing or idling (away from keyboard), uploading a large amount of low-effort content, and mass-friending large numbers of users. As the VRChat official puts it, “You gain these ranks simply by playing VRChat – as you explore worlds, make friends, and create content, you will gain more trust, which determines our Trust Rank.”

In addition to these standard ranks, the TS also includes special ranks like “Nuisance” and “VRChat Team.” The Nuisance rank is assigned as a form of penalty for users who engage in misconduct and typically results in being blocked by default. This means they cannot be seen or heard unless other users choose to unblock and unmute them manually. The VRChat Team rank is reserved exclusively for members of the VRChat team and is identifiable by a special tag. Each trust rank is associated with a specific color displayed on the user’s nameplate. This visual indicator helps users quickly identify the trust level of others in their vicinity. For example, Known Users might have orange nameplates, while Trusted Users have purple ones. Because of that, many use colors to describe user groups, such as orange as an alternative name for the Known User. The nameplates are a vital aspect of the user interface, providing immediate context about who one interacts with. With increasing trust ranks come greater privileges. For instance, users ranked Visitor and below are restricted from uploading content, while those above this rank are granted this ability. Trust and Known Users have the unique option to present themselves as ordinary Users, influencing how the system and other participants interact with them.

A set of safety settings is often used in combination with TS, which allows users to configure how users in each rank are treated regarding their visibility. As shown by Table 2, including voice, avatar, avatar audio, custom animations, shaders, particles, and lights, users can enable or disable other users’ features based on their ranks. Users can configure each rank with a unique setting, referring to Fig. 1. For example, if one disables all the features of users in Visitor Rank, then with the setting, all Visitor users will be invisible to the user. If the voice setting is activated, then one can hear the voice of users in these ranks, but not other

---

<sup>9</sup> <https://docs.VRChat.com/docs/VRChat-safety-and-trust-system>





**Figure 1.** A screenshot of the Safety Settings interface, illustrating how users can manage the visibility of each trust rank. Upon selecting a specific rank, the interface reveals various features that can be enabled



## “People are Way too Obsessed with Rank”: Trust System in Social...

**Table 2.** Individual safety setting

Name	Function
Voice	Mutes or unmutes a user’s microphone (voice chat)
Avatar	Hiding or showing a user’s avatar and all avatar features
User Icon	Controlling user icon visibility for Trust Rank
Avatar Audio	Enabling or disabling sound effects from a user’s avatar (not their microphone)
Shaders	When shaders are disabled, all shaders on a user’s avatar are reverted to Standard
Animations	Enabling or disabling custom animations on a user’s avatar
Light & Particles	Enabling or disabling particle systems on a user’s avatar, as well as any light sources

things. By default, all safety settings are disabled for trusted users, meaning their avatar effects are fully visible unless manually adjusted by other users.

It also allows for hiding or showing specific users. For example, if a user encounters toxic users with ranks that the user does not set restrictions on, the user can choose to block those specific toxic users. The safety system will disable all the features of the users’ avatars.

### 3 Method

#### 3.1 Data collection

To answer our research questions, we utilized the discussion data from the “r/VRChat” subreddit, the largest online community for VRChat users, with more than 148 thousand users at the time of writing. Researchers frequently browse the subreddit as they are interested in the social VR platform. The online discussions contain the naturally and directly shared experiences of VRChat users. They are different from conducting interviews or surveys, which heavily rely on self-reports and require participants to recount, which may be limited by recall bias (Gorin and Stone 2001) and social desirability bias (Fisher 1993). Thus, we deemed collecting and analyzing discussion data appropriate as the first endeavor.

We used Reddit’s API, which allows us to fetch data by relevant keywords. The API enables us to collect all threads containing keywords in either content, user comments, or titles. Our initial set of keywords was carefully chosen based on a thorough review of VRChat’s official documentation<sup>10</sup> on the TS, combined with the first author’s extensive experience on the platform, spanning over three years, including “known user, visitor, nuisance, trust user, trust system.” To enhance the scope and depth of our dataset, we further analyzed the first hundred threads from our initial data collection. This detailed review was crucial in identifying

<sup>10</sup> <https://docs.vrchat.com/docs/vrchat-safety-and-trust-system>

additional keywords frequently used by users in their discussions about the TS. It was observed that users often used terms like “rank system” interchangeably with the TS and “trust ranks” as a general term for different levels within the system. Recognizing the significance of these terms in the context of user discussions, we incorporated “trust rank” and “rank system” as complementary additional keywords in our data collection process. This expansion of keywords, based on actual user language and discussion patterns, ensured that our data collection was not only thorough but also aligned with the users’ perspectives and terminologies. It provided us with a more nuanced and authentic dataset, reflecting the honest conversations and opinions of VRChat users regarding the TS.

We scraped all the relevant published data in r/VRChat, about 11596 posts in total (including threads and comments). The dates of the posts ranged from 2018 to 2022. The data scraped included the title, body, URL of each thread, and the total number and content of the associated comments, as well as their metadata like timestamp and upvotes. After the data scraping, we embarked on refining our dataset. Our primary focus was on eliminating irrelevant data that did not pertain to the TS. For example, while many posts mentioned keywords, not all of them were relevant to discussions about the TS; for example, “The loud music from the participants’ avatars became a nuisance, disrupting the event I had planned.” Such posts were identified and removed from our dataset. Additionally, we encountered replicated data due to the keyword-based approach of our data retrieval. This occurred because some threads touched upon multiple keywords, leading to their multiple inclusions in our dataset. To address these challenges, we thoroughly reviewed all 647 threads. Our goal was to identify and eliminate posts that, despite including keywords, did not discuss the TS and duplicated threads and associated comments. Through this rigorous filtering process, we refined our dataset to 6,028 posts, of which 362 were threads and the remainder were comments. This final dataset was considered representative and relevant for our analysis, focusing on user interactions and opinions regarding the TS in VRChat.

### 3.2 Data analysis

We aimed to gain a deep understanding with nuanced details about users’ perception of the TS in VRChat. We employed a qualitative analysis followed by previous work with a similar dataset (Ma and Kou 2021). We adopted the inductive thematic analysis method from (Braun and Clarke 2006) to code the collected data. Two coders met regularly throughout the entire analysis phase. We first immersed ourselves in reading the data to familiarize ourselves with the whole dataset. After the familiarization, two researchers separately returned to the dataset to generate initial codes for words, sentences, or paragraphs for the same hundred posts, including comments. Then, the coders met to discuss the initial list of codes. We went over and compared our codes. We identified differences where

## “People are Way too Obsessed with Rank”: Trust System in Social...

one coder might have a code not identified by another. In such cases, we would discuss whether the code was sufficiently distinctive in its semantics or whether it could be grouped into other codes. Through this process, we formed a consensus among the codes. Cohen’s Kappa ( $\kappa$ ) was used to measure inter-rater reliability (IRR) (McDonald et al. 2019; Viera et al. 2005). IRR on the transcript was  $\kappa=0.80$  ( $SD=0.20$ ), with codes ranging from strong to nearly perfect agreement. Then, we split the rest of the data to code. We employed iterative coding, initially generating codes in a data-driven manner. We coded segments of text based on their inherent meaning. For example, a user comment such as “The ranks damage first impressions, which kills the desire to interact.” was coded as ‘damage motivation to interact’. Throughout the analysis, we constantly moved between the codes and the associated data, collaboratively synthesizing these codes to develop higher-level themes. For instance, specific codes like ‘damage motivation to interact’, ‘blocking based on ranks helps segregation’, and ‘muting based on ranks helps segregation,’ were collectively grouped under a broader theme, ‘inhibit interactions among users from different ranks’. When a new code appeared, we compared it with our existing codes to either integrate it into previous themes or create a new theme for it. Our iterative process culminated in the formation of overarching themes that encapsulated the complex user perceptions of the TS in VRChat.

Overall, standard ethical guidelines are followed (Franzke et al. 2020; GDPR. Eu. n.d.) to collect user data and perform analysis. We consulted with our institution’s ethics committee on potential ethical concerns. The data are openly accessible and have a public viewing expectation. Additionally, they do not involve sensitive information. The gathering and processing are regarded as no more than minimal risk to individuals. Therefore, we do not need to request ethics approval for our study. But, we are aware that the HCI community has several reservations regarding the use of publicly accessible data (Adams 2022). Fiesler suggests that these concerns should be framed within “the broader context of the benefits of scientific discovery” (Fiesler 2019). The implications are that researchers should be thoughtful and sensitive in their interactions with the data and in their purposes for using it. In this research, we utilized multiple approaches to protect the people involved in our study. We rephrased our quotes to reduce the researchability for tracking back to the original posts. We eliminated all information that could be used to identify an individual. The researchers in charge are the only ones who have access to the data, which have been safely saved on our password-protected devices. Researchers in the team also had meetings to explore the possible negative effects and potential positive outcomes of this research in addition to the aforementioned measurements. We reasoned that: 1) The research provides an understanding of the TS used in the most popular social VR platform, VRChat, exploring how users perceive it and weighing its pros and cons. The outcomes can provide implications for similar systems to better cope with harassment. 2) The research can act as a vehicle to understand harassment in social VR platforms.

## 4 Findings

The findings reveal a complex interplay of user logic, system implementation, and the resultant social dynamics, leading to varied interpretations and reactions among the user base. The TS aims to assess user trustworthiness based on engagement and behavior. Many users believe that, by logic, longer, non-toxic participation enhances their trust rank, while toxic actions may negatively impact it. However, the system's lack of transparency and perceived inconsistencies in rank assignment have led to distrust among users. Trust ranks often indicate a user's experience and engagement with the platform. However, this categorization fosters stereotypes and inhibits interactions between users of different ranks, leading to social segregation. Additionally, the system inadvertently promotes discrimination and conflict, particularly against lower-ranked or new users. Furthermore, the TS introduces an element of competitiveness into the social platform, shifting focus from social interaction to rank achievement. Overall, while the TS is based on logical principles, its implementation and user interpretations have resulted in complex social dynamics, including issues of trust, discrimination, and competitiveness.

### 4.1 Believe in the system: Logic-driven trust

A fundamental function of TS is to evaluate and present a user's trustworthiness. A portion of the VRChat community places their trust in the TS, valuing its logicbased approach to evaluating user trustworthiness. Ideally, the longer users spend on the platform, the higher their rank becomes. However, if they tend to engage in toxic behavior towards other users, they are more likely to be reported, blocked, or votekicked, which many users believe can negatively impact trust ranks (VRChat does not confirm the influence of these actions on user ranks.). Thus, toxic users cannot upgrade their ranks, so users with high ranks tend not to be obnoxious. A user's comment illustrates the view,

“The trust is based on how long you've been using the platform without being obnoxious to others. You are either new or you're not... I believe that if someone has played for a long time without becoming a nuisance. People will trust them.”

The quote shows that people who tend to be toxic are less likely to have high ranks. They spend more time on the platform while upgrading their trust rank, indicating that they conduct a limited amount of harmful behaviors.

Many users regarded the TS acts as a form of deterrence. In order to get high ranks, users need to put time and effort into the platform. This investment creates a disincentive for high-level users to engage in harassment or toxic behavior, as

“People are Way too Obsessed with Rank”: Trust System in Social...

once they are being massively blocked, vote-kicked, or reported, resulting in losing trust level or even a ban on the account. As demonstrated by a user,

“Upgrading to a higher rank is difficult and requires a lot of time and effort from users. So I think high-level users seldom do behaviors that may result in harmful consequences to their ranks, such as behaviors that may lead to reports. But low-level users don’t have this concern.”

According to the user’s comment, higher-ranked individuals are typically more mindful of their actions as they have invested significant time and effort in attaining their rank. They are less likely to do things that lead to negative consequences in their ranks. Conversely, those who tend to exhibit toxic behavior are most likely with lower ranks, such as the Visitors.

#### 4.2 Inappropriate implementation results in distrust

Many users believe in the system because of its logic, but its inappropriate implementation results in distrust from other users. The algorithm’s opacity is a contributing factor. Users are left in the dark about how their ‘trustworthiness’ is calculated, aside from the system’s reliance on metrics such as time spent on VRChat, number of friends, and contributions to the platform. Many users assumed the ranks could only demonstrate how a user perform in the above factors but could not tell whether a user is trustworthy or whether they are more or less likely to carry out toxic behavior. A user’s comment encapsulates a common sentiment,

“The Trust System has never been a reliable approach for evaluating g people. Since it solely indicates user experience and their contribution to the game asset.”

The TS is perceived more as a measure of user activity than an accurate reflection of trustworthiness. It fails to delve into the nuances of user interactions, leaving many to speculate and doubt its effectiveness in identifying toxic behavior. It is echoed by another viewer: *“I am positive that the ranking system is purely RNG. No one truly knows how to move up”*. This quote demonstrates users’ distrust due to the system’s opacity. RNG stands for a random number generator, which describes things that produce random results in games.

Another reason is due to the precarity of the algorithm. Many users observed that individuals with better performance in each factor, such as more time, more contributions, more friends, and fewer violations of the community code, received a lower trust rank than others. For example,

“This rating system is so erratic... I am still a new user with 65 hours, half of my friends with 40 are known users or higher, and I have four avatars posted.”

Such experiences highlight the perceived inconsistencies in the ranking process. Users who believe they perform better in key metrics still lag in rank compared to others, leading to questions about the system’s fairness and accuracy.

Additionally, the user community has observed that high rank does not necessarily equate to high moral standards or non-toxic behavior. For example,

“Don’t trust people based on their ranks. I once reasoned that I might loosen my security setting when around “known” and “trusted” users. This was foolish on my side and ignorant. Morality is not ranked in the ranking system. It’s just a way to gauge how much time someone has invested in the game. People who have played the game longer are more likely to be aware of its safety flaws. Some utilize it as knowledge to protect themselves. But I’ve had known and trusted users crash me.”

The above quote is from a user who wrote to warn others that Known and Trusted users are also toxic and should set Safety Settings against them. Their deep understanding of the system’s intricacies could be used for protection or, conversely, to exploit less informed users.

#### 4.3 Consider trust ranks as references to user activities

As mentioned in the previous section, the trust ranks are perceived to be able to demonstrate how users perform in the factors the TS uses to calculate their ranks, such as time spent. In practice, users utilize this characteristic of trust ranks to interpret other users’ information. For example, they regard the ranks as an indication of a user’s familiarity and experience with the platform, which can help in finding suitable friends. The following observation reflects a perception among users,

“The users with higher ranks are more likely to be familiar with or comfortable with VRChat culture and not find things so strange, which is the only thing the ranks tell me.”

As stated by the quote, the user thought the primary information that a trust rank can provide is a user’s familiarity with the platform. Trust ranks become a shorthand for differentiating users familiar and comfortable with the VRChat culture from those newcomers who are easily surprised by things on VRChat.

“People are Way too Obsessed with Rank”: Trust System in Social...

Additionally, the rank can also serve as a valuable tool to discern users' interest on the platform, distinguishing between those who are deeply committed and those who are more transient in nature. Users who attain higher ranks typically demonstrate prolonged and active involvement, signaling a robust interest in the platform. Consequently, many individuals utilize these ranks as a means to connect with others who share a similar level of investment and enthusiasm, preferring to interact with those who are not just briefly exploring out of curiosity, as noted by a user,

“And when it comes to elitism, I do it... with their rank. Because that informs me of the level of users' interest in the social game. I don't like to build relationships with people I know that they will become bored after a few weeks and then I won't be able to see them. That's why I want to make sure the folks I know show symptoms of enjoying the game.”

The use of trust ranks as a barometer for users' commitment and longevity on the platform is evident in this user's statement. High-ranking users are often seen as more engaged and less likely to abandon the platform on a whim, making them preferable companions for those seeking long-term interactions.

#### 4.4 Give rise to stereotypes and inhibiting interaction across ranks

While intended to assess and display user trustworthiness, the system's categorization process leads to unintended social dynamics. Trust ranks classify users into different groups, which provides the basis for stereotype formation. In addition, these ranks can hinder interaction among users with varying ranks.

##### 4.4.1 *Provide the fertile land for stereotypes*

Trust ranks, by dividing users into distinct categories, lay the groundwork for stereotype formation. Each rank becomes more than just a measure of trustworthiness; it evolves into a label with certain generalizations about the users. With a rank, everyone represents their specific rank. One user observed,

“Everyone in the said group is responsible for the reputation of their group. A few members of group A do a few things that groups B and C don't like, and now, all of a sudden, everyone in group A is marked because a few members of group A, who have no interaction with anyone else from the group, did something. People have actually said, and I quote: ‘The trusted users are the least trustable’. ‘Oh, they are simply visitor ranks; let's just block them.’ ”



The quote reflects how users often judge others based on the actions of a few individuals within the same rank category. Such generalizations lead to sweeping judgments about entire groups, irrespective of individual differences.

Stereotyping cuts across all ranks. Higher-ranked users are sometimes perceived as overly invested or lacking a life outside the virtual world. Conversely, lower-ranked users are labeled as annoying, toxic, and squeaky. A user described the experience from the point of users in high rank,

“Some people will think of you as one of the high-rank stereotypes. Like an asshole, a nolifer, someone who knows everything there is to know about the game, and so on. And many trusted/known users do not want to deal with that nonsense. So they set their safety setting up in the hopes of avoiding it.”

High-rank users are assumed to spend most of their time in VRChat; thus, they have no real life. In response to these stereotypes, users often adapt their settings and interactions to shield themselves from perceived negative aspects associated with other rank.

#### 4.4.2 *Inhibit interactions among users from different ranks*

Due to the formed stereotypes, many users observed that the ranks impact users' first impressions. It creates invisible barriers discouraging interactions between users of differing ranks. This segregation, fueled by preconceived notions about each rank's characteristics, limits the opportunities for diverse social encounters. Additionally, the safety settings allowing users to limit the visibility of certain ranks exacerbate segregation. For example,

“It is a social game that is strongly reliant on first impressions, and If you remove the first impression, there is no stimulus to interact. Sure, it's good to have the choice to disable someone's avatar, but nobody actually asked for further segregation based on some arbitrary numbers that aren't even consistent between checks.”

The trust ranks hamper the interaction among users. The first impression of people is ruined due to the displayed ranks, which also alleviates the stimulus to interact among users. Although the user acknowledged that the avatar disabling function (referring to Table 2) is useful, it also further segregates users of different ranks.

#### 4.5 *Inducing discrimination and conflicts particularly against low-rank users*

The TS, designed to foster a safe and respectful online environment, paradoxically gives rise to new forms of discrimination and conflict, particularly

“People are Way too Obsessed with Rank”: Trust System in Social...

affecting users with lower ranks. This unintended consequence stems from the hierarchical nature of the system, where different levels of trust create divisions and tensions within the community.

#### 4.5.1 *Inducing discrimination*

It has been observed that trust ranks can cause discrimination, especially toward the low rank. There are clearly distinct inferiority and superiority emerging in the ranks, with the Trust User being the most superior and owning diverse privileges. The hierarchical structure causes unfair treatment toward the low-rank compared to their higher-ranking counterparts. A user who was being demoted describes his experience, “*You get treated noticeably worse when you’re lower rank.*” He observed a distinct difference in attitudes towards him compared to when he held the higher rank. The discrimination can create significant challenges for newcomers. They appear with a Visitor rank, which is the lowest trust rank. The experience shared by a user in the lowest rank, ‘Visitor’, illustrates this point,

“It undoubtedly adds more discriminating action against newer gamers. I normally don’t have problems with people since I don’t see or feel the need to be toxic (it is, after all, a social game), but I was blocked by two different people merely for having a grey rank.”

The user was blocked simply because of their low rank, contributing to the feeling that the newcomers were targeted with discrimination. With the implementation of the TS, the platform may not be newcomer-friendly.

#### 4.5.2 *Generating conflicts*

Beyond discrimination, the TS’s hierarchical structure also fosters direct conflicts. High-rank users may form cliques and engage in exclusionary or even hostile behaviors towards those with lower ranks. As stated by a user,

“I saw a group of green and orange ranked members would circle people with grey names and belligerently yell at them and harass them until they were blocked or the target left the server.”

This account depicts a scenario where users with higher ranks use their status to intimidate and harass lower-ranked individuals. Such actions create a hostile environment, contradicting the TS’s objective of promoting a safe and respectful community.

Additionally, the system’s mechanics, such as blocking and vote-kicking, are sometimes exploited to manipulate ranks, leading to unjust demotions and furthering

the divide between users. As shown in a comment, “*Friends of mine get demoted in trust by people with mass blocking to demote user ranks*”. Many users in VRChat generally believe that blocking and vote-kicking are being calculated in the system, negatively impacting users’ trust ranks. The safety instruments are supposed to be used to reduce trolling, but here, they are utilized as tools for malicious purposes.

#### 4.6 Making a social platform more competitive

The implementation of trust ranks inadvertently shifts the environment towards competitiveness. High-rank users have more privileges, and ranks are dynamic and changeable, leading to a competitive atmosphere. A user illustrated the view,

“There’s literally no reason to display ”rank” ingame and the safety features should be split up in Users and Friends. VRChat is a social platform, not a competitive esport game.”

This sentiment expressed by a user underscores the discomfort with the visible hierarchy created by the trust ranks. The public display of ranks transforms what is meant to be a social, interactive space into a field of competition where users are constantly aware of their status relative to others. It is echoed by other users, for example,

“People are way too obsessed with rank these days. Just play the game, make some friends, and learn a little bit about yourself.” Replied by another user, “Because getting a higher rank has both tangible and intangible benefits.”

In the first quote, the users noticed people are over-obsessed with ranks. It would benefit individuals to focus more on the social aspects of the platform, including expanding their social circle and self-discovery. As shown previously, to upgrade to a higher trust level, users dedicated excessive time on the platform, some exceeding one thousand hours, learning the unknown algorithm behind the TS and uploading more than a hundred models, such as worlds and avatars. The second quote revealed that the competition is partially due to the tangible and intangible benefits a higher rank can bring to users.

## 5 Discussion

We set out to investigate how users perceive the TS in VRChat by gathering and analyzing data from one of the largest VRChat online forums, “r/VRChat. Our research revealed that many users considered the system logically valid, but its

## “People are Way too Obsessed with Rank”: Trust System in Social...

inappropriate implementation leads to distrust. In addition, this system seems to promote negativity and intensify competition among users. Next, we will discuss our findings further and give implications from our findings for better designing such systems.

### 5.1 Trust system imposes challenges for newcomers

Newcomers play a crucial role in community growth as a way to replenish the people who leave the community (Kraut and Resnick 2012). Additionally, newcomers often bring fresh ideas and content to a platform. However, attracting and integrating them into an existing community is challenging (Kiene et al. 2016; Zhu et al. 2014). Newcomers do not yet have the same level of commitment to a group as experienced members. Therefore, they are extremely sensitive to their own early experiences on a new platform. Even minor difficulties could cause them to abandon a community altogether (Arguello et al. 2006; Schweik and English 2012). Given that social VR is an emerging online space, it is essential to attract and integrate newcomers for its continued development. However, the TS presents various challenges for newcomers. The TS's hierarchical structure, distinguished by varying ranks, inherently fosters a power dynamic that can lead to discrimination or harmful interactions between established members and newcomers. In this sense, TS mimics the symbolic system and affords symbolic power to high-rank users (Bourdieu 1979). It is deeply ingrained in the attitudes and behaviors of these high-rank users, reinforcing the hierarchical structure that benefits and perpetuates the dominance of those in established positions of power (Cattani et al. 2014).

*Trust System Hinders Newcomers' Adaption to the Novel Online Space.* The TS hinders the adaption of newcomers to the platform in several ways. First, the TS is often used alongside safety settings, which allow users to control the visibility of other users according to their ranks. A low trust rank here implies that users with the rank tend to exhibit toxic behavior or have not proven trustworthy. As a result, many highrank users often avoid contacting low-rank users by either refraining from engaging with them or directly employing safety settings to mute their voices or even making them completely invisible. Such avoidance can be harmful to newcomers and to the whole community, as it limits newcomers' communication with other users in the novel online spaces, especially the high-rank and experienced users. Newcomers may act in ways that insult other users or otherwise undermine the community's ability to function due to a lack of familiarity with the community norms that govern behaviors (Kraut and Resnick 2012). Social VR is a novel online space that most people have never been in contact with or have no idea of, and its norms differ from other online spaces (Blackwell et al. 2019; Maloney et al. 2020). Therefore, such avoidance hinders the learning process and leads newcomers to easily make mistakes and

be punished. Second, it assigns the newcomers the lowest rank with their nameplates. The opacity and precarity of the system also create obstacles for newcomers to upgrade and discourage their participation. Compared to high-rank users, newcomers are unfamiliar with the algorithm. They need to learn how it works. Learning a new algorithm takes time and requires additional labor (Ma and Kou 2021). Lack of access and guidance from experienced users can hinder this learning, thus limiting the adaptation of newcomers.

*The High-rank Oppresses and Discriminates Low-rank Users.* The TS makes it more difficult for newcomers to adapt to the new community. It also facilitates the oppression of newcomers from old-timers. The ranks form a hierarchy among users with clear inferiority and superiority. The Trusted User are the most superior ones with privileges from the system, while newcomers are the Visitor at the bottom of the ladder. The TS implicitly indicates that low-rank users are toxic or unreliable. Therefore, other users often assume that newcomers are toxic or have been demoted or punished by the platform, even though they just arrived (Cai and Wohn 2021). Such a stereotype is not conducive for newcomers to adapt to the new platform, even leading to group attack to and dropout of newcomers. Furthermore, it is commonly expressed that high-rank users are generally treated much better than low-rank users.

At the same time, many high-rank users build their competitive advantages over newcomers by leveraging the opacity of the TS. High-rank users are more informed about how the system works, knowing how and what actions make users upgrade and downgrade. Instead of helping new users, many high-rank users leverage their algorithmic knowledge to suppress these newcomers. Our findings show that the high-rank maliciously demote low-rank users through activities such as blocking (blocking is generally considered harmful to a user's rank). In some cases, high-rank users even coordinated as groups to expel low-rank users and suppress newcomers by massively blocking them.

Our study contributes to the previous literature in using the reputation system as a disciplinary system for addressing toxicity. We have found that ranks in the reputation system may impede the growth of newcomers in an online social community as the system may facilitate disconnection between newcomers and experienced users. Meanwhile, the formed hierarchy due to the assigned ranks may cause discrimination and even harassment against newcomers. Furthermore, the findings supplement the research on social VR harassment. Previous research on the prevalence of toxicity in social VR considers different reasons, such as the affordances of VR, the demographic of the typical VR user (Blackwell et al. 2019), and finally, the influence of gaming culture (Blackwell et al. 2019; Freeman et al. 2022a). Our findings show that the toxicity in social VR is also partially caused by design choices. It also evidences and expands Whitney Phillips' argument (Phillips 2015) that people are not the only cause for toxicity online but also system designs, further into social VR. TS also appears to enhance the

## “People are Way too Obsessed with Rank”: Trust System in Social...

tension between “game” and “social” culture, making it more difficult to find a consensus on appropriate norms given the diversity in the community. We will discuss it in the next section.

### 5.2 Trust ranks provoke and intensify competitiveness

Due to the gaming culture’s influence and novelty of social VR, social VR users may interpret these online spaces in various ways, such as games, social media platforms, or others. As shown in our study, users refer to VRChat in different ways, including “game,” “social game,” and “social platform.” Trust ranks may enhance social VR users’ impression that social VR is a competitive game, thus hindering the formation of less toxic norms.

Social VR is an online space that facilitates an immersive and embodied social experience for its users. Individuals engage in various social activities in social VR, from dancing, sleeping, and partying to fulfilling their different social needs (Maloney and Freeman 2020). Social VR is heavily influenced by gaming culture (Blackwell et al. 2019). For example, many social VR platforms integrate competitive games into their platforms as part of their content, such as “Among Us<sup>11</sup>” and “Prison Escape!<sup>12</sup>” the very popular game worlds on VRChat that require social VR users to complete designated tasks and compete with others. In addition, many of them, like VRChat and Recroom, are on the Steam Game Store for download. These factors inevitably make some users think that social VR platforms are related to games instead of only for socializing, as evidenced by the different ways viewers call them in our study. For individuals who perceive social VR as a form of a game instead of a social platform toxicity and conflicts are acceptable. Game players tend to accept toxicity, such as trolling, profanity, etc., and think that it is an inextricable part of game culture (Beres et al. 2021; Adinolf and Turkay 2018), thus behaving so. The two groups of people with different perceptions of social VR may create tensions. Many users urge and warn that the platform is meant for social use instead of competition (e.g., “*VRChat is a social platform, not a f\*\*king competitive esport game.*”). The rank design, commonly used in games, strengthens the platform’s competitiveness and enhances users’ impression that the platform is a game. The increased competitiveness due to ranks, can make it difficult and unbearable for users who come to this platform to socialize. As our findings demonstrate, many users complain that ranks make VRChat a game and cause too many users to be obsessed with the ranks instead of socializing.

In addition, the increased competitiveness may exacerbate the harassment in the platform as competitiveness fostering toxicity (Adinolf and Turkay 2018; Kou 2020, Gaikwad et al. 2016). Previous research argues that the flux of new

---

<sup>11</sup> <https://vrchat.com/home/launch?worldId=wrlddd036610-a246-4f52-bf01-9d7cea3405d7>

<sup>12</sup> <https://vrchat.com/home/launch?worldId=wrlD14750dd6-26a1-4edb-ae67-cac5bcd9ed6a>

users may encourage the reconstruction of the toxic social norms in social VR (Shriram and Schwartz 2017). However, the premise is that more newcomers consider those platforms as social spaces instead of competitive esports. Designs like TS that encourage competition through ranks can strengthen the view that the platform is for gaming instead of socializing and, therefore, support game-oriented norms with a high tolerance for toxicity.

### 5.3 Implications for future designs

Based on our findings, it appears that many users have faith in TS because they perceive its logic in managing toxicity as reasonable. Specifically, if a user consistently displays toxic behavior towards others, then the more time they spend on the platform, the more likely they are to be reported, blocked, and vote-kicked (those users generally believe the actions have negative impacts on their ranks). Thus, toxic users cannot upgrade their rank. But at the same time, the inappropriate implementation of the system breeds doubt about its effectiveness. Our findings evidence and expand Whitney Phinllips' argument (Phillips 2015) that people are not the only cause for toxicity online but also system designs. In this study, we argue that when integrating a design into a novel context, we should be cautious of the negative consequences of design choices. Based on the challenges we identified in the TS, we provide several suggestions. Since the VR environment is highly immersive and embedded, we also aligned the design implications with real-world cases to show their potential to improve the TS design. By addressing these implications, our goal is to improve the functionality and fairness of the TS, fostering a more inclusive and supportive environment within the social VR community.

#### 5.3.1 *Improving the system's credibility among users*

One issue raised by the study in implementation is the negative effect of missing transparency of the instrument. There is no specific information available on how the ranks are computed, except that they are related to users' time on the platform, the number of friends, and contributions. The opacity of algorithms may result in the breakdown of the system's credibility (Kizilcec 2016; Eslami et al. 2019). Adding some dimensions of transparency can help address these users' concerns (Ma and Kou 2023). However, the level of transparency must be carefully designed, as it can complicate the user interaction with the system and provide opportunities for malicious users to game the system (Eslami et al. 2019).

In addition, developers can actively build and maintain trust between the system and users. For example, to set up communication channels and communicate with users in a timely manner. Another effective strategy is to involve users in the algorithm system and allow them to participate in the user-human-AI moderation loop (Schulenberg et al. 2023). For example, after recognizing the user's prosocial behavior, the system can let community members evaluate whether



“People are Way too Obsessed with Rank”: Trust System in Social...

the data or behavior of a user warrants an upgrade. This approach increases the user’s influence in the decision-making process, enhancing trust, as suggested by (Smith et al. 2020) (which aims to create a reputation system to grant the editing right). Similar to real-world cases such as trust and confidence toward legitimate jurisdiction systems, the platform should improve transparency in the decision-making process to improve users’ perceived procedural justice and fairness to improve the credibility of the TS (Wallace and Goodman-Delahanty 2021).

### 5.3.2 *Diversing the factors that the algorithm uses*

The TS is designed to evaluate individuals’ trustworthiness to reinforce prosocial behaviors and discourage antisocial behaviors. Therefore, it is reasonable to include factors related to their antisocial behaviors, such as their history of harmful actions (e.g., hate speech and harassment) and more prosocial activities, such as helping others. Balancing the opposite factors related to content moderation can mitigate the algorithmic bias toward newcomers and make the ranks more closely aligned with the function of this system. Recently, great progress has been made in using machine learning to detect whether the behavior of online users is beneficial to the community (Bao et al. 2021). The system may consider using machine learning as a way to directly detect the quality of users’ behaviors to establish the records. But at present, this method is only for voice and text-based content and does not account for other modalities common in social VR, such as body language, which can also harm users in the novel space. It shows the need for more research on recognizing the quality of nonverbal actions.

### 5.3.3 *Enhancing newcomer integration*

Facilitating the integration of newcomers into online social spaces is essential for building successful online communities (Kraut and Resnick 2012). This is particularly relevant for newcomers in social VR, who often encounter challenges. Drawing inspiration from the structured approach used in France for new drivers<sup>13</sup> and Wikipedia’s Welcoming Committee<sup>14</sup>, we can devise strategies to address the “newbies problem”, integrating these with the trust system to foster a supportive environment.

Firstly, a structured orientation and theory learning process similar to that of new drivers in France can be implemented. This would involve a tutorial system for new users that is linked to their TS. Completing different stages of this orientation could enhance their trust score, ensuring that they are knowledgeable about the platform’s norms and functionalities. Following the theoretical orientation, practical lessons and guided experiences, akin to driving lessons, could be introduced. New users could be paired with experienced mentors who would

---

<sup>13</sup> <https://www.expatica.com/fr/living/transportation/french-drivers-license-107659/>

<sup>14</sup> [https://en.wikipedia.org/wiki/Wikipedia:Please\\_do\\_not\\_bite\\_the\\_newcomers](https://en.wikipedia.org/wiki/Wikipedia:Please_do_not_bite_the_newcomers)

guide them through various aspects of the VR world, such as social etiquette, interaction techniques, and feature usage. This mentorship could be a voluntary program where experienced users opt-in to assist newbies, earning trust points for their active guidance and support.

In addition, the high-rank users here tend to be unfriendly new users, as we discovered. The TS might create designs to prompt experienced users, especially those with high ranks, of their responsibilities, taking a cue from Wikipedia's approach of encouraging seasoned editors to be kind to newcomers. High-ranking users in VRChat could receive periodic reminders about the importance of being welcoming and supportive. Positive interactions with newcomers could lead to an increase in the trust score of the experienced user. Lastly, encouraging continuous learning and community engagement is vital. This can be facilitated through regular events, workshops, and discussions. Similar to how drivers are expected to stay updated with road rules and safety measures, VRChat users could be encouraged to keep abreast of community standards, new features, and best practices.

By adopting these methods, newcomers can more effectively be integrated into the community, and be well-informed, comfortable, and engaged from the outset. This strategy not only improves the experience for newcomers and cultivates a more welcoming, inclusive, and responsible community.

#### *5.3.4 Designing to provide an alternative for the visibility of the rank*

We discussed above that the ranks in the system bring negatives in the social setting, such as conflicts among users. Furthermore, from another point of view, this system can be compared to public shaming as a means of requiring users to abide by the norms. The ranks communicate not only users' roles but also their power and authority in the community, similar to offline settings such as policing (Loader 1997). The ranks serve as a basis for this shaming, which has been observed in various internet cultures to enforce particular behavioral norms or role expectations (Klonick 2015). To mitigate the negative effects caused by ranks, we suggest providing an alternative for the visibility of the rank, such as allowing users to set their ranks as private. In this way, although this system does not have the function of exposing users to the public gaze. But it still has a disciplinary function, which is self-discipline (Foucault 1995). Private ranks can be seen as a form of self-discipline, which reminds users that they have been in the disciplinary gaze of the platform (which establishes their ranks). And users can also be part of this gaze, observing their ranks to see their level and whether they are toxic. Such behavior can be due to the person's pride and self-esteem (Tomkinson and Van Den Ende 2022). Beyond using ranks to foster prosocial behaviors, other methods exist, such as providing rewards to users when their reputation ranks upgrade. For example, the Honor System in League of Legends rewards players who advance in the honor

“People are Way too Obsessed with Rank”: Trust System in Social...

level with items that can be found in loot boxes (League of Legends Wiki. [n.d](#)). These incentives can induce users to achieve a high trust level, thus encouraging good behaviors while preventing conflicts and discrimination due to ranks.

## 6 Limitations and future work

There are several limitations. First, the research data collected includes many negative comments on the TS. It is possible that these comments are influenced by certain user patterns, such as feeling undervalued or believing that other closely connected users are undervalued. However, we lack information on the background of the users who post these comments, including their ranks. This limits our ability to fully understand the context behind their opinions. Second, our study also did not attempt to gauge the proportion of different opinions of users. Obtaining data on the proportion of these opinions would give researchers a more comprehensive understanding of the TS. Additionally, it is essential to address the potential bias introduced by our method of selecting keywords for the initial data collection. We chose our keywords based on VRChat’s official documentation and our research experiences. This approach, while grounded in a thorough review and practical knowledge, may carry the risk of subjective bias.

Several potential research directions also emerge from the study. First, research on how users perceive and use safety tools to mitigate toxicity, such as safety bubbles and blocking is necessary. Users may use these tools differently than developers expect, as our research demonstrated. A previous study on the flagging behaviors of users in a game setting reported that these behaviors are highly related to the competitive gaming context. There are differences between what the platform wants users to flag and what users think should be flagged (Kou and Gui 2021). The social VR context could condition the way users use these tools. Second, another potential research direction is the discrimination issue in social VR. While analyzing the data, the researchers found that many factors may cause discrimination. Trust rank is only one type among them. We want to understand the discrimination issue on social VR and explore the difference between the issue on social VR and that on conventional social networks such as Facebook. Lastly, to comprehensively assess the effectiveness of VRChat’s TS in mitigating toxicity, it’s crucial to extend the investigation beyond VRChat. This involves comparing VRChat’s TS with similar systems on other digital platforms. By exploring and contrasting user perceptions and impacts of these systems across different platforms, insights can be gained into how their design and implementation influence user behavior and community engagement. Through such an examination, the effectiveness of these systems in promoting positive interactions, managing behavior, and reducing toxicity can be evaluated. This broader analysis not only sheds light on VRChat’s system and offers valuable perspectives on reputation system design and management in the digital realm.

## 7 Conclusion

Our study shows that users perceive the system differently. Many users believe in it because its logic of coping with toxicity is reasonable. However, the problems in its implementation, such as opacity and inconsistency, cause distrust. The trust ranks, while intended to promote positive interactions, build the foundation based on which users form stereotypes and discourage communication between users of different ranks. In addition, the ranks introduce a hierarchy among users, which fuels discrimination and conflicts, particularly against the low rank. We further discuss that this discouragement of communication and discrimination against low-rank is particularly harmful to newcomers, as they come to the platform with the lowest rank at the bottom of the hierarchy. Newcomers are the source of the growth of an online community, particularly for emerging platforms like social VR. Trust ranks present challenges to newcomers, which may impede community growth. Furthermore, the ranks contribute to a competitive atmosphere that hinders the formation of less toxic norms. Finally, we provide concrete design implications for similar systems in the future.

## Acknowledgements

We thank the anonymous reviewers for their constructive comments, which significantly strengthened the paper. This work was supported by the Research Council of Finland (\#357270).

**Author contributions** Qijia Chen wrote the manuscript and collected and analyzed the data. Jie Cai structured and revised the manuscript critically. Giulio Jacucci revised the manuscript and was involved in data analysis.

**Funding** Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). Research Council of Finland (\#357270).

**Data availability** The collected discussion data that support the findings of this study are available on request from the corresponding author.

## Declarations

**Competing interests** The authors declare no competing interests.

**Ethical Approval** Local regulations do not require formal ethics review.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

“People are Way too Obsessed with Rank”: Trust System in Social...

the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams, Norman Nicholas. 2022. ‘Scraping’ Reddit Posts for Academic Research? Addressing Some Blurred Lines of Consent in Growing Internet-Based Research Trend during the Time of Covid-19. *International Journal of Social Research Methodology* 27 (1): 47–62. <https://doi.org/10.1080/13645579.2022.2111816>.
- Adinolf, Sonam, and Selen Turkey. 2018. “Toxic Behaviors in Esports Games: Player Perceptions and Coping Strategies.” *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, Melbourne, Australia*, October, 365–372. New York: ACM Press. <https://doi.org/10.1145/3270316.3271545>.
- Arguello, Jaime, Brian S. Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Rosé, and Xiaoqing Wang. 2006. “Talk to Me.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Canada*, April, 959–968. New York: ACM Press. <https://doi.org/10.1145/1124772.1124916>.
- Bao, Jiajun, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and David Jurgens. 2021. “Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations.” *Proceedings of the Web Conference 2021, Ljubljana, Slovenia*, April, 1134–1145. New York: ACM Press. <https://doi.org/10.1145/3442381.3450122>.
- Barreda-Ángeles, Miguel, and Tilo Hartmann. 2022. Psychological Benefits of Using Social Virtual Reality Platforms during the Covid-19 Pandemic: The Role of Social and Spatial Presence. *Computers in Human Behavior* 127 (February): 107047. <https://doi.org/10.1016/j.chb.2021.107047>.
- Beres, Nicole A., Julian Frommel, Elizabeth Reid, Regan L. Mandryk, and Madison Klarkowski. 2021. “Don’t You Know That You’re Toxic: Normalization of Toxicity in Online Gaming.” *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan*, May, 1–15. New York: ACM. <https://doi.org/10.1145/3411764.3445157>.
- Blackwell, Lindsay, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. “Harassment in Social Virtual Reality : Challenges for Platform Governance.” *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–25. <https://doi.org/10.1145/3359202>.
- Bourdieu, P. 1979. Symbolic Power. *Critique of Anthropology* 4 (13–14): 77–85. <https://doi.org/10.1177/0308275x7900401307>.
- Braun, Virginia, and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3 (2): 77–101. <https://doi.org/10.1191/1478088706qp0630a>.
- Cai, Jie, and Donghee Yvette Wohn. 2021. After Violation but before Sanction: Understanding Volunteer Moderators’ Profiling Processes toward Violators in Live Streaming Communities. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–25. <https://doi.org/10.1145/3479554>.
- Cai, Jie, Sagnik Chowdhury, Hongyang Zhou, and Donghee Yvette Wohn. 2023. Hate Raids on Twitch: Understanding Real-Time Human-Bot Coordinated Attacks in Live Streaming Communities. *Proceedings of the ACM on Human-Computer Interaction* 7 (CSCW2): 1–28. <https://doi.org/10.1145/3610191>.
- Cai, Jie and Donghee Yvette Wohn. 2023. “Understanding Moderators’ Conflict and Conflict Management Strategies with Streamers in Live Streaming Communities.” *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany*, April, 1–12. New York: ACM. <https://doi.org/10.1145/3544548.3580982>.

- Cai, Jie, Donghee Yvette Wohn, and Mashaël Almoqbel. 2021. "Moderation Visibility: Mapping the Strategies of Volunteer Moderators in Live Streaming Micro Communities." *ACM International Conference on Interactive Media Experiences, Virtual Event*, June, 61–72. New York: ACM. <https://doi.org/10.1145/3452918.3458796>.
- Cattani, Gino, Simone Ferriani, and Paul D. Allison. 2014. Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields. *American Sociological Review* 79 (2): 258–281. <https://doi.org/10.1177/0003122414520960>.
- Chandrasekharan, Eshwar, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can't Stay Here. *Proceedings of the ACM on Human-Computer Interaction* 1 (CSCW): 1–22. <https://doi.org/10.1145/3134666>.
- De Simone, Francesca, Jie Li, Henrique Galvan Debarba, Abdallah El Ali, Simon N. B. Gunkel, and Pablo Cesar. 2019. "Watching Videos Together in Social Virtual Reality: An Experimental Study on User's Qoe." *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, March, 890–891. New York: ACM. <https://doi.org/10.1109/vr.2019.8798264>.
- Despotovic, Zoran, and Karl Aberer. 2004. Possibilities for managing trust in P2P networks. *EPFL Technical Report* 1–41. <https://infoscience.epfl.ch/record/52675>. Accessed 22 Mar 2024.
- Dosono, Bryan, and Bryan Semaan. 2019. "Moderation Practices as Emotional Labor in Sustaining Online Communities." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK*, May, 1–13. New York: ACM. <https://doi.org/10.1145/3290605.3300372>.
- Eslami, Motahhare, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. "User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK*, May, 1–14. New York: ACM. <https://doi.org/10.1145/3290605.3300724>.
- Fiesler, Casey. 2019. Ethical considerations for research involving (speculative) public data. *Proceedings of the ACM on Human-Computer Interaction* 3 (GROUP): 1–13. <https://doi.org/10.1145/3370271>.
- Fisher, Robert J. 1993. Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20 (2): 303–315. <https://doi.org/10.1086/209351>.
- Foucault, Michel. 1995. *Discipline and Punish: The Birth of the Prison*, 2nd edn. New York: Vintage Books.
- Franzke, Aline Shakti, Anja Bechmann, Charles Melvin Ess, and Michael Zimmer. 2020. Internet research: Ethical guidelines 3.0. *The International Association of Internet Researchers* 4(1): 1–83. <https://aoir.org/reports/ethics3.pdf>. Accessed 22 Mar 2024.
- Freeman, Guo, Divine Maloney, Dane Acena, and Catherine Barwulor. 2022a. "(Re)Discovering the Physical Body Online: Strategies and Challenges to Approach Non-Cisgender Identity in Social Virtual Reality." *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, New Orleans, USA, April, 1–15. New York: ACM. <https://doi.org/10.1145/3491102.3502082>.
- Freeman, Guo, Samaneh Zamanifard, Divine Maloney, and Dane Acena. 2022b. Disturbing the Peace: Experiencing and Mitigating Emerging Harassment in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW1): 1–30. <https://doi.org/10.1145/3512932>.
- Frenkel, Sheera, and Kellen Browning. 2021. The Metaverse's dark side: Here come harassment and assaults. *The New York Times*. <https://www.nytimes.com/2021/12/30/technology/metaverse-harassment-assaults.html>. Accessed 22 Mar 2024.
- Gaikwad, Snehal Kumar (Neil), Durim Morina, Adam Ginzberg, Catherine Mullings, Shirish Goyal, Dilrukshi Gamage, Christopher Diemert, et al. 2016. "Boomerang: Rebounding the Consequences of Reputation Feedback on Crowdsourcing Platforms." *Proceedings of the 29th*

## “People are Way too Obsessed with Rank”: Trust System in Social...

- Annual Symposium on User Interface Software and Technology, Victoria, Canada, October, 625–637.* New York: ACM. <https://doi.org/10.1145/2984511.2984542>.
- GDPR.Eu. n.d. “General Data Protection Regulation (GDPR) Compliance Guidelines.” <https://gdpr.eu/>. Accessed 14 March 2024.
- Gorin, Amy A., and Arthur A. Stone. 2001. Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. *Handbook of Health Psychology* 23: 405–413.
- Grimmelmann, James. 2015. The Virtues of Moderation. *Yale Journal of Law and Technology* 17 (1): 68.
- Hendrikx, Ferry, Kris Bubendorfer, and Ryan Chard. 2015. Reputation Systems: A Survey and Taxonomy. *Journal of Parallel and Distributed Computing* 75 (January): 184–197. <https://doi.org/10.1016/j.jpdc.2014.08.004>.
- Jhaver, Shagun, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. ‘did You Suspect the Post Would Be Removed?’ *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–33. <https://doi.org/10.1145/3359294>.
- Jiang, Jialun Aaron, Charles Kiene, Skyler Middler, Jed R. Brubaker, and Casey Fiesler. 2019. Moderation Challenges in Voice-Based Online Communities on Discord. *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–23. <https://doi.org/10.1145/3359157>.
- Jonas, Marcel, Steven Said, Daniel Yu, Chris Aiello, Nicholas Furlo, and Douglas Zytko. 2019. “Towards a Taxonomy of Social VR Application Design.” *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, Barcelona, Spain, October*, 437–444. New York: ACM. <https://doi.org/10.1145/3341215.3356271>.
- Jøsang, Audun, Roslan Ismail, and Colin Boyd. 2007. A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems* 43 (2): 618–644. <https://doi.org/10.1016/j.dss.2005.05.019>.
- Jøsang, Audun. 2006. “Trust and Reputation Systems.” *Foundations of Security Analysis and Design IV*, 209–245. [https://doi.org/10.1007/978-3-540-74810-6\\_8](https://doi.org/10.1007/978-3-540-74810-6_8).
- Kiene, Charles, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. “Surviving an ‘Eternal September.’” *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, USA, May*, 1152–56. New York: ACM. <https://doi.org/10.1145/2858036.2858356>.
- Kilteni, Konstantina, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21 (4): 373–387. [https://doi.org/10.1162/pres\\_a\\_00124](https://doi.org/10.1162/pres_a_00124).
- Kim, Soomin, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1): 1–26. <https://doi.org/10.1145/3449161>.
- Kizilcec, René F. 2016. “How Much Information? Effects of Transparency on Trust in an Algorithmic Interface.” *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, USA, May*, 2390–2395. New York: ACM. <https://doi.org/10.1145/2858036.2858402>.
- Klonick, Kate. 2015. “Re-Shaming the Debate: Social Norms, Shame, and Regulation in an Internet Age.” *SSRN Electronic Journal*, 1029. <https://doi.org/10.2139/ssrn.2638693>.
- Kolesnichenko, Anya, Joshua McVeigh-Schultz, and Katherine Isbister. 2019. “Understanding Emerging Design Practices for Avatar Systems in the Commercial Social VR Ecology.” *Proceedings of the 2019 on Designing Interactive Systems Conference, San Diego, USA, June*, 241–252. New York: ACM. <https://doi.org/10.1145/3322276.3322352>.
- Kou, Yubo. 2021. Punishment and Its Discontents: An Analysis of Permanent Ban in an Online Game Community. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–21. <https://doi.org/10.1145/3476075>.



- Kou, Yubo, and Xinning Gui. 2021. "Flag and Flaggability in Automated Moderation." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan*, May, 1–12. <https://doi.org/10.1145/3411764.3445279>.
- Kou, Yubo. 2020. "Toxic Behaviors in Team-Based Competitive Gaming: The Case of League of Legends." *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Virtual Event, Canada*, November, 81–92. New York: ACM. <https://doi.org/10.1145/3410404.3414243>.
- Kraut, Robert E., and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. Cambridge, MA: Mit Press.
- League of Legends Wiki. n.d. "Honor." *League of Legends Wiki*. Fandom, Inc. <https://leagueoflegends.fandom.com/wiki/Honor>. Accessed 14 March 2024.
- Li, Jie, Yiping Kong, Thomas Röggl, Francesca De Simoni, Swamy Ananthanarayan, Huib de Ridder, Abdallah El Ali, and Pablo Cesar. 2019. "Measuring and Understanding Photo Sharing Experiences in Social Virtual Reality." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK*, May, 1–14. New York: ACM. <https://doi.org/10.1145/3290605.3300897>.
- Loader, Ian. 1997. Policing and the Social: Questions of Symbolic Power. *The British Journal of Sociology* 48 (1): 1–18. <https://doi.org/10.2307/591907>.
- Ma, Renkai, and Yubo Kou. 2021. 'how Advertiser-Friendly Is My Video?': YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–25. <https://doi.org/10.1145/3479573>.
- Ma, Renkai, and Yubo Kou. 2023. 'defaulting to Boilerplate Answers, They Didn't Engage in a Genuine Conversation': Dimensions of Transparency Design in Creator Moderation. *Proceedings of the ACM on Human-Computer Interaction* 7 (CSCW1): 1–26. <https://doi.org/10.1145/3579477>.
- Maloney, Divine, Guo Freeman, and Donghee Yvette Wohn. 2020. 'Talking without a Voice' Understanding Non-Verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2): 1–25. <https://doi.org/10.1145/3415246>.
- Maloney, Divine, and Guo Freeman. 2020. "Falling Asleep Together: What Makes Activities in Social Virtual Reality Meaningful to Users." *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Virtual Event, Canada*, November, 510–521. New York: ACM. <https://doi.org/10.1145/3410404.3414266>.
- Maselli, Antonella, and Mel Slater. 2013. The Building Blocks of the Full Body Ownership Illusion. *Frontiers in Human Neuroscience* 7: 83. <https://doi.org/10.3389/fnhum.2013.00083>.
- McDonald, Nora, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research. *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–23. <https://doi.org/10.1145/3359174>.
- McVeigh-Schultz, Joshua, Elena Márquez Segura, Nick Merrill, and Katherine Isbister. 2018. "What's It Mean to 'Be Social' in VR? Mapping the Social vr Design Ecology." *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems, Hong Kong, China*, May, 289–294. New York: ACM. <https://doi.org/10.1145/3197391.3205451>.
- McVeigh-Schultz, Joshua, Anya Kolesnichenko, and Katherine Isbister. 2019. "Shaping Pro-Social Interaction in VR: An Emerging Design Framework." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK*, May, 1–12. <https://doi.org/10.1145/3290605.3300794>.
- Pan, Christina A., Sahil Yakhmi, Tara P. Iyer, Evan Strasnick, Amy X. Zhang, and Michael S. Bernstein. 2022. Comparing the Perceived Legitimacy of Content Moderation Processes: Contractors, Algorithms, Expert Panels, and Digital Juries. *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW1): 1–31. <https://doi.org/10.1145/3512929>.

## “People are Way too Obsessed with Rank”: Trust System in Social...

- Phillips, Whitney. 2015. *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge, MA: Mit Press. <https://doi.org/10.7551/mitpress/10288.001.0001>.
- Piitulainen, Roosa, Perttu Hämäläinen, and Elisa D Mekler. 2022. “Vibing Together: Dance Experiences in Social Virtual Reality.” *CHI Conference on Human Factors in Computing Systems, New Orleans, USA*, April, 1–18. New York: ACM. <https://doi.org/10.1145/3491102.3501828>.
- Rachel, Metz. 2022. *Harassment Is a Problem in VR, and It's Likely to Get Worse*. <https://www.cnn.com/2022/05/05/tech/virtual-reality-harassment/index.html>. Accessed 22 Mar 2024.
- Resnick, Paul, and Richard Zeckhauser. n.d. Trust among strangers in internet transactions: Empirical analysis of Ebay's reputation system. *The Economics of the Internet and E-Commerce* 127–157. [https://doi.org/10.1016/s0278-0984\(02\)11030-3](https://doi.org/10.1016/s0278-0984(02)11030-3).
- Ruohomaa, Sini, Lea Kutvonen, and Eleni Koutrouli. 2007. “Reputation Management Survey.” *The Second International Conference on Availability, Reliability and Security (ARES'07)*, Vienna, Austria, April 103–11. New York: IEEE. <https://doi.org/10.1109/ares.2007.123>.
- Sabri, Nazanin, Bella Chen, Annabelle Teoh, Steven P. Dow, Kristen Vaccaro, and Mai Elsherief. 2023. “Challenges of Moderating Social Virtual Reality.” *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany*, April, 1–20. New York: ACM. <https://doi.org/10.1145/3544548.3581329>.
- Saffo, David, Caglar Yildirim, Sara Di Bartolomeo, and Cody Dunne. 2020. “Crowdsourcing Virtual Reality Experiments Using VRChat.” *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, USA*, April, 1–8. New York: ACM. <https://doi.org/10.1145/3334480.3382829>.
- Schoenebeck, Sarita, Carol F Scott, Emma Grace Hurley, Tammy Chang, and Ellen Selkie. 2021. Youth trust in social media companies and expectations of justice: Accountability and repair after online harassment. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1): 1–18. <https://doi.org/10.1145/3449076>.
- Schulenberg, Kelsea, Guo Freeman, Lingyuan Li, and Catherine Barwulor. 2023. ‘Creepy Towards My Avatar Body, Creepy Towards My Body’: How Women Experience and Manage Harassment Risks in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 7 (CSCW2): 1–29. <https://doi.org/10.1145/3610027>.
- Schweik, Charles M., and Robert C. English. 2012. “The Importance of Open-Source Software Commons.” *Internet Success*, June, 2–12. <https://doi.org/10.7551/mitpress/9780262017251.003.0001>.
- Seering, Joseph, Robert Kraut, and Laura Dabbish. 2017. “Shaping pro and Anti-Social Behavior on Twitch through Moderation and Example-Setting.” *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, USA*, February, 111–125. New York: ACM. <https://doi.org/10.1145/2998181.2998277>.
- Seering, Joseph, Tianmi Fang, Luca Damasco, Mianhong “Cherie” Chen, Likang Sun, and Geoff Kaufman. 2019. “Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors.” *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK*, May, 1–14. New York: ACM. <https://doi.org/10.1145/3290605.3300836>.
- Shriram, Ketaki, and Raz Schwartz. 2017. “All Are Welcome: Using VR Ethnography to Explore Harassment Behavior in Immersive Social Virtual Reality.” *2017 IEEE Virtual Reality (VR)*, Los Angeles, USA, 225–226. New York: IEEE. <https://doi.org/10.1109/vr.2017.7892258>.
- Slater, Mel, Beau Lotto, M Arnold, and María V. Sánchez-Vives. 2009. How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario De Psicología* 40 (2): 193–210. <http://diposit.ub.edu/dspace/bitstream/2445/49643/1/631349.pdf>. Accessed 22 Mar 2024.

- Smith, C. Estelle, Bowen Yu, Anjali Srivastava, Aaron Halfaker, Loren Terveen, and Haiyi Zhu. 2020. “Keeping Community in the Loop: Understanding Wikipedia Stakeholder Values for Machine Learning-Based Systems.” *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, USA*, April, 1–14. New York: ACM. <https://doi.org/10.1145/3313831.3376783>.
- Srinivasan, Kumar Bhargav, Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Chenhao Tan. 2019. Content Removal as a Moderation Strategy: Compliance and Other Outcomes in the Changemyview Community. *Proceedings of the ACM on Human-Computer Interaction 3* (CSCW): 1–21. <https://doi.org/10.1145/3359265>.
- Steiger, Miriah, Timir J Bharucha, Sukrit Venkatagiri, Martin J. Riedl, and Matthew Lease. 2021. “The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support.” *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan*, May, 1–14. New York: ACM. <https://doi.org/10.1145/3411764.3445092>.
- Sykownik, Philipp, Linda Graf, Christoph Zils, and Maic Masuch. 2021. “The Most Social Platform Ever? A Survey about Activities & Motives of Social VR Users.” *2021 IEEE Virtual Reality and 3D User Interfaces (VR), Virtual Event*, March, 546–554. New York: IEEE. <https://doi.org/10.1109/vr50410.2021.00079>.
- Sykownik, Philipp, Divine Maloney, Guo Freeman, and Maic Masuch. 2022. “Something Personal from the Metaverse: Goals, Topics, and Contextual Factors of Self-Disclosure in Commercial Social VR.” *CHI Conference on Human Factors in Computing Systems, New Orleans, USA*, April, 1–17. New York: ACM. <https://doi.org/10.1145/3491102.3502008>.
- Tomkinson, Sian, and Benn van den Ende. 2021. ‘Thank You for Your Compliance’: Overwatch as a Disciplinary System. *Games and Culture 17* (2): 198–218. <https://doi.org/10.1177/15554120211026257>.
- Uttarapong, Jirassaya, Jie Cai, and Donghee Yvette Wohn. 2021. “Harassment Experiences of Women and LGBTQ Live Streamers and How They Handled Negativity.” *ACM International Conference on Interactive Media Experiences, Virtual Event, USA*, June, 7–19. New York: ACM. <https://doi.org/10.1145/3452918.3458794>.
- Vidal, Alfonso. 2023. Genshin impact reputation guide. *Codashop Blog SG*. <https://news.codashop.com/sg/genshin-impact-reputation-system/>. Accessed 22 Mar 2024.
- Viera, Anthony J., Joanne M. Garrett, et al. 2005. Understanding interobserver agreement: The kappa statistic. *Fam med 37* (5): 360–363. <https://pubmed.ncbi.nlm.nih.gov/15883903>. Accessed 22 Mar 2024.
- Wallace, Anne, and Jane Goodman-Delahunty. 2021. Measuring Trust and Confidence in Courts. *International Journal for Court Administration 12* (3): 1–17. <https://doi.org/10.36745/ijca.418>.
- Wang, Tzu-Yang, Yuji Sato, Mai Otsuki, Hideaki Kuzuoka, and Yusuke Suzuki. 2019. “Effect of Full Body Avatar in Augmented Reality Remote Collaboration.” *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan*, March, 1221–1222. New York: IEEE. <https://doi.org/10.1109/vr.2019.8798044>.
- Xiao, Sijia, Shagun Jhaver, and Niloufar Salehi. 2023. Addressing Interpersonal Harm in Online Gaming Communities: The Opportunities and Challenges for a Restorative Justice Approach. *ACM Transactions on Computer-Human Interaction 30* (6): 1–36. <https://doi.org/10.1145/3603625>.
- Xiao, Sijia, Coye Cheshire, and Niloufar Salehi. 2022. “Sensemaking, Support, Safety, Retribution, Transformation: A Restorative Justice Approach to Understanding Adolescents’ Needs for Addressing Online Harm.” *CHI Conference on Human Factors in Computing Systems, New Orleans, USA*, April, 1–15. New York: ACM. <https://doi.org/10.1145/3491102.3517614>.
- Zheng, Qingxiao, Xu. Shengyang, Lingqing Wang, Yiliu Tang, Rohan C. Salvi, Guo Freeman, and Yun Huang. 2023. Understanding Safety Risks and Safety Design in Social VR Environments.

“People are Way too Obsessed with Rank”: Trust System in Social...

*Proceedings of the ACM on Human-Computer Interaction* 7 (CSCW1): 1–37. <https://doi.org/10.1145/3579630>.

Zhu, Haiyi, Jilin Chen, Tara Matthews, Aditya Pal, Hernan Badenes, and Robert E. Kraut. 2014. “Selecting an Effective Niche: An Ecological View of the Success of Online Communities.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada*, April, 301–310. New York: ACM. <https://doi.org/10.1145/2556288.2557348>.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.