12-2017

Managing Online Toxic Disinhibition: The Impact of Identity and Social Presence

Hyerin Kim
University of Texas at Austin, Hyerin.Kim@utexas.edu

Younghoon Chang
United International College, younghoonc@uic.edu.hk

Managing Online Toxic Disinhibition: The Impact of Identity and Social Presence

Hyerin Kim
University of Texas at Austin
Hyerin.Kim@utexas.edu

Younghoon Chang
United International College
younghoonc@uic.edu.hk

ABSTRACT

Drawing on psychology literature in the intersection of impression management and social presence, we examine the potential of identity-communicating artifacts for managing toxic online disinhibition. Multiplayer online games are chosen as the research context where online disinhibition behavior is pervasive at all times. Gamer history and synchronous voice chat are suggested to construct and communicate identity. We hypothesize the influence of these IT artifacts on social presence, enjoyment, and toxic online disinhibition. By analyzing primary survey and participatory observation data, we validated the hypotheses. The results verified the promising role of identity-communicating IT features in managing toxic online behaviors. The paper proposes a mechanism for understanding underlying online gamers' behavior and provides implications for designing multi-player games and managing online gaming communities.

Keywords

Online disinhibition, Social presence, Online identity, Synchronous communication, Multi-player online game

INTRODUCTION

Multi-player online games have been growing remarkably popular, and the online gaming world provides another place where people interact with each other. The anonymous environment and competitive nature of the online games have caused some side effect called toxic disinhibition, which includes trolling, cheating, profanity, and cyberbullying (Blackburn & Kwak, 2014). Online toxic disinhibition behavior negatively affects individual users' online service experience and the success of the online services. In reaction to this, online services experiencing user toxic disinhibition as well as academics have tried to investigate this unpleasant social phenomenon and to find ways to manage it. Multi-player online games are one of the extreme cases where online toxic disinhibition is present at all times, and the managers of such online services are always looking for ways to solve this problem (Maher, 2016).

Earlier research tried to verify the factors involved with the psychological mechanism under online disinhibition (Suler, 2004). Some previous studies have argued that online disinhibition is inevitable due to anonymous and invisible nature of the online setting. Little research, however, has suggested mechanisms that can possibly reduce online toxic disinhibition. We, therefore, examine the potential of identity-communicating IT artifacts for managing toxic online disinhibition in multi-player online games, drawing on psychology literature intersecting impression management and social presence. We argue that the improvement of identity perpetuation and communication synchronicity is likely to lower online toxic disinhibition by stimulating impression management motivation.

LITERATURE REVIEW

Online Disinhibition Effect

The online disinhibition effect is defined as the tendency for people to feel unrestrained in online settings and therefore conduct themselves differently than the way they would in offline settings (Suler, 2004). This online disinhibition is closely related to the specific/particular nature and traits of online media, such as dissociative anonymity, invisibility, asynchronicity, solipsistic introjections, dissociative imagination, and minimization of status and authority (Suler, 2004; Zhao, Grasmuck, & Martin, 2008). These unique characteristics/components of online environments distort individual users' perception of existence and presence and hinder the ability of individuals to construct and communicate their significant identity. The notion of dissociation due to the mediated settings of online environments is essential to comprehend/ facilitate the comprehension of online disinhibition. Online users are likely to disassociate themselves from any recurring identities or hold only temporal identities which they consider negligible. Accordingly, toxic disinhibitions commonly including profanity and vulgarity have often flourished as if the users do not care of their impression to others or how they are evaluated. Thus, we set our main independent variables to examine if IT artifacts can complement the dissociative nature of online environments. We believe that providing gamers with game elements that improve or alternate communications mediated by technology will develop a different sense of their online social presence, which will consequently reduce the occurrence of toxic disinhibition.

Social Presence Theory

Social Presence Theory (SPT) explains how communication medium can shape the way people communicate by affecting interactants' social presence (Short, Williams, & Christie, 1976). Social presence is defined as "the ability of participants in a community of inquiry to project themselves socially and emotionally, as 'real' people (i.e., their full personality), through the medium of communication being used." (Garrison, Anderson, & Archer, 2000, p. 94). That is to say, perceived social presence is about the degree to which one person feels to be present with another person in a mediated setting. Research on social presence and technology-mediated interaction has emphasized the importance of individual perceptions of social presence over the objective traits of a communication medium (Lowenthal, 2010). According to SPT, when interactants recognize each other as well as themselves as a "real" person in the setting, there will be effective and satisfactory interpersonal involvement and exchange (Gunawardena & Zittle, 1997). SPT, therefore, provides a critical clue to comprehend and manage disinhibition effect due to the dissociative online setting.

RESEARCH MODEL

We develop a research model and propose seven hypotheses as presented in Figure 1.

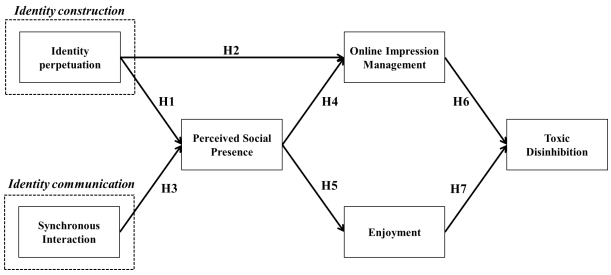

Figure 2. Results of Hypotheses Tests

Hypotheses

As suggested in H1 and H2 in Figure 1, we hypothesize that IT artifacts that develop one's sense of online identity are likely to enhance perceived social presence. The mediation of technology has created challenges in establishing one's identity, for the technology mediation results in social cue deficiencies and cessation of managing individual identity. In addition, in many times, by dissociative anonymity and asynchronicity, online user identity seems to be discontinued (Yee & Bailenson, 2007). Online environments tend to be dissociative in terms of one's identity as argued in prior research (Suler, 2004). Dissociative anonymity makes individuals difficult to associate themselves or others to certain labels to distinguish each other; invisibility dissociates individuals from physical entities; minimization of authority dissociates individuals from the status or power that the individuals hold.

We hypothesize that game elements can help identity construction and maintenance so that one feels that their identity perpetuates in the online game. Deep profiling and persistent labeling are examples of IT artifacts supporting identity communication (Ma & Agarwal, 2007). In our research context, persistent labeling such as game ID plays a single name to present oneself and the play history search makes their identity continued. Hence, we posited H1.

H1: When one feels that their identity is perpetuating, he/she is likely to perceive higher social presence.

When people obtain a sense of continuity of their identity, they desire more to be understood who they are and fundamentally get motivated to manage how they will be seen and evaluated by others (Swann et al., 2004). Moreover, the sense an identity and impressions around it will be stuck together and perpetuate affects one's motivation to manage their impression.

H2: When one feels that their identity is perpetuating, he/she is likely to have higher intention to manage their online impression.

We also hypothesize that synchronous interaction by using new media and communication tools with their unique affordances can enhance social cues of the existence of both self and others and help identity communication (Mennecke, Triplett, Hassall, & Heer, 2011). Social presence theory, by definition, elaborates the degree of perceived salience between two communicators of being existent on a communication medium. Additionally, the synchronous interaction will promote intimacy and immediacy, which are maintained to be two main concepts that primarily consist of social presence as a construct (Short et al., 1976). Therefore, synchronous interaction by using in-game voice chat will increase one's perception of social presence.

H3: When one experiences more synchronous interaction, he/she is likely to perceive higher social presence.

In the earlier work of Goffman (1956) about the presentation of self in everyday life, it was suggested in a social setting when we enter the presence of others, our behaviors are aimed at impression management. Perceived social presence is directly related to the way of online interaction (Tu & McIsaac, 2002). Research has examined that when individuals perceive a kind of social situation, people tend to follow the rules of conducts between them or act based on the fact that others might monitor and even evaluate them (Walther, 2007). Hence, higher perceived social presence will lead to one's impression management.

H4: When one perceives higher social presence, he/she is likely to have higher intention to manage their online impression.

Consistent benefits of social presence have been identified and agreed upon. Particularly, research has

shown the strong link between high perceived social presence and enjoyment. Particularly, in the hedonic online gaming setting, enjoyment is a significant part of user satisfaction. Gunawardena and Zittle (1997) verified that social presence is a significant predictor of interactants' satisfaction within a computer-mediated form of communication, explaining about 60% of the variance. Therefore,

H5: When one perceives higher social presence, he/she is likely to enjoy.

Physical dissociation such as invisibility has been studied to be one of the critical factors causing toxic online disinhibition. In their experiment research, Lapidot-Lefler and Barak (2012) asserted that factors related to online social communication factors, particularly lack of eye-contact, are main causes of the manifestation of toxic online disinhibition. Thus, by increasing the degree to which people perceive to be socially existent, toxic disinhibition can be managed and reduced. Furthermore, in self and social evaluative point of view, when they have a motivation to manage their desired self-impression, they are more likely to restrains extreme toxic behaviors. Accordingly, we hypothesized that

H6: When one has higher intention to manage their online impression, they are likely to conduct less toxic disinhibition behaviors.

Psychologically toxic disinhibition is closely related to restrained negative feelings such as anger, shame, agony or wrath. When people are enjoying and feel pleasant, they do not have negative emotions to express on less restraint settings, which include online services or communities. A number of studies also have shown that satisfaction promotes pleasant participation (Givertz & Segrin 2005). Therefore,

H7: When one enjoys more, they are likely to conduct less toxic disinhibition behaviors.

RESEARCH METHODOLOGY

Research Focus

We invited users playing Overwatch, an online team-based multiplayer first-person shooter game, and users playing League of Legend (LoL), a multiplayer online battle arena (MOBA) game. Both games were proclaimed the world's most played online game in terms of the number of hours that users accessed and played. In Overwatch, synchronous voice chat is included inside the game to synergize team driven actions. In LoL, there is no such voice chat function, but some gamers use external Voice over IP (VoIP) services such as TalkOn or Skype.

Data Collection and Sample

The research involved participatory observations followed by paper-based self-reports. Due to the possible social desirability bias of respondents in their self-report about their toxic behaviors, we combined participatory observation along with the survey. We conducted the research at a PC bang, which is a local version of an Internet café mostly associated with online gaming in South Korea. PC bangs are equipped with relatively standardized high-quality computers designed to support gaming, online game licenses on the computers, and broadband Internet access for an hourly fee around $1 (Huhh, 2008; Hjorth, 2011). The atmosphere of PC bang makes the participants more focused on playing games than on participating in the research, and the standardized environment reduces the effect of possible external factors.

Our sample consisted of 180 Overwatch players and 211 LoL players. After removing incomplete data using listwise deletion, the final sample included 144 players and 180 players respectively. 72.09% participants were male and 55.2% participants reported being in the 20 to 29 age range. The participants were given the instant 1-hour PC bang vouchers upon their completion to fill in the questionnaire.

Operationalization

A survey questionnaire was developed by reviewing the measurement items from the previous literature. Some measurements were modified to be more appropriate for the context of online gaming and toxic disinhibition. Identity perpetuation (IDP) was measured with the degree of the use of player history and log function as a proxy. The use of synchronous in-game voice chat was used as the proxy to measure synchronous interaction (SYNC). Perceived social presence (PSP) measurements were brought from Gunawardena and Zittle (1997). The measures for online impression management (IMP) are adapted from prior literature studying impression management in broadly defined online community setting (Kim, Chan, & Atreyi, 2012; Riemer & Shavitt, 2011). Enjoyment (ENJ) items were adapted from Duffy, Shaw, and Stark (2000). Finally, we applied four toxic disinhibition (TOX) items developed by Udris (2014) and our raters checked the toxic disinhibition level they heard in participatory observation over one battle. The questionnaire employed the seven-point Likert scale.

PRELIMINARY RESULT

Our collected data were examined using Partial Least Square (PLS). We used SPSS 18.0 and Smart PLS 2.0.M3 to analyze the data. PLS differentiates between a measurement and structural model, so we examined two models separately in two-stage. In this manuscript of research in progress, we focused on analyzing the first sample data of 49 LoL gamers.

Measurement Model

The assessment of the measurement model includes the estimation of internal consistency for reliability and the tests of convergent and discriminant validity. Table 1

shows the indices for reliability and convergent validity. It shows that our measurement model demonstrated good reliability and convergent validity (Bollen, 1989).

	# of Items	Cronbach 's alpha	CR	AVE
IDP	3	0.72	0.84	0.63
SYNC	4	0.74	0.83	0.56
PSP	5	0.89	0.92	0.71
IMP	3	0.84	0.90	0.75
ENJ	3	0.70	0.82	0.61
TOX	2	0.83	0.92	0.86

Table 1. Measurement Model

Structural Model

Figure 2 demonstrates the results of the structural model. Overall the model explained 57.5 percent of the variance in toxic online disinhibition. Impression management motivation negatively affected toxic disinhibition and the hypothesis was statistically significantly supported. However, the path coefficient of enjoyment explaining toxic disinhibition was not significant. The path coefficients around the perceived social presence were revealed to be all significant. Synchronous interaction was the strongest predictor of perceived social presence with the path coefficient of 0.828.

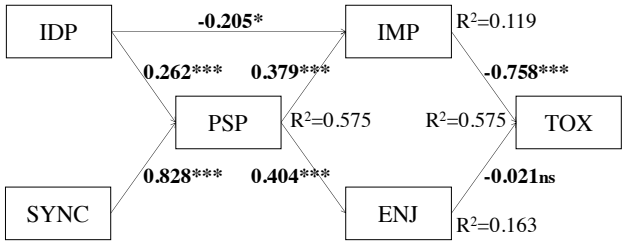

Figure 2. Results of Hypotheses Tests

Our results indicated that identity perpetuation negatively impacted online impression management with the path coefficient of -0.205, which was significant at $p<0.05$ level. The result was opposite from our hypothesis that when people feel that their identity is perpetuating, they are likely to have higher intention to manage their online impression.

DISCUSSION

The first motivation of this paper is to propose ways to manage toxic online behavior, which hinders user experience and the success of online services or communities over time. Theoretically, we aimed to comprehend the mechanisms underlying dissociative nature of online media and disinhibition effect by centering on the perceived social presence in technology-mediated settings.

Multiplayer online games are chosen to be the research context where online disinhibition behavior prevails and users always complain about the disturbing experience. We set gamer history and synchronous voice chat to be identity constructing and communicating artifacts and designed the survey and observation to examine the effect of those. We hypothesized that those identity developing artifacts would enhance perceived social presence, subsequently, increase impression management intention and enjoyment, and, consequently, negatively impact toxic disinhibition. The results verified the promising role of identity-communicating IT features in managing toxic online behaviors.

Contributions

This research presents a theoretical avenue to link IT affordance and user behavior management. We introduced perceived social presence as the main psychological mechanism through which ICT artifacts helps reduce online toxic disinhibition. Using social presence theory as a basis, this research has modeled what makes people start to manage their impression to others online.

IT artifacts such as Gamer's play log and synchronous voice chat were studied in relation to online identity construction, maintenance, and communication. Those artifacts increase the sense of social presence. Thus, this research contributes by explaining IT affordance in regards to online identity.

From a practical perspective, this study can provide some hints to solve perplexing and problematic online user behaviors or, at least, to reduce those behaviors to a certain extent.

Limitations and Future Research

Although our findings confirm our hypotheses on toxic behavior, we note that there are limitations to be considered. Those game elements were suggested to develop one's online identity and social presence, but the fact remains that our research premised these competitive real-time game setting. Applying those same elements to other domains or games with different characteristics need to be done cautiously. For example, in some situations where they do not need real-time discussion and coordination, such identity perpetuation and synchronicity could cause one's fatigue.

Future Plan

First of all, we will examine the variables that we eliminated for the simple model. There are a couple of hypotheses that are turned out to be insignificant. This might due to the elimination of control variables such as game self-efficacy, gender, age, and of additional

underlying variables such as subjective norm or self-esteem. Secondly, we will review the relevant theories more in details and revise our models based on the propositions or ideas we have missed. Third, we will develop competing models to find the model that can have more implications. Finally, as mentioned in the methodology section, we originally conducted the study on gamers of two different game communities. We will complete the analysis on both dataset and perform path comparisons to increase generalization of our argument.

REFERENCES

1. Blackburn, J., & Kwak, H. (2014). STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games. In WWW' 14.
2. Bollen, K. A. (1989). Structural Equations with Latent Variables. New York: John Wiley & Sons.
3. Duffy, M. K., Shaw, J. D., & Stark, E. M. (2000). Performance and satisfaction in conflicted interdependent groups: When and how does self-esteem make a difference?. Academy of Management Journal, 43(4), 772-782.
4. Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. The Internet and Higher Education, 2(2-3), 87-105.
5. Givertz, M., & Segrin, C. (2005). Explaining personal and constraint commitment in close relationships: The role of satisfaction, conflict responses, and relational bond. Socio Personal Relationship, 22(6), 757-775.
6. Goffman, E. (1959). The Presentation of Self in Every Day Life. New York: Doubleday.
7. Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. American Journal of Distance Education, 11(3), 8–26.
8. Hjorth, L. (2011). Games and Gaming: An Introduction to New Media. London: Bloomsbury.
9. Huhh, J.-S. (2008). Culture and Business of PC Bangs in Korea. Games and Culture, 3(1), 26–37.
10. Kim, H., Chan, H. C., & Atreyi, K. (2012). What Motivates People to Purchase Digital Items on Virtual Community Websites? The Desire for Online Self-Presentation. Information Systems Research, 23(4), 1232–1245.
11. Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. Computers in Human Behavior, 28(2), 434–443.
12. Lowenthal, P. R. (2010). The evolution and influence of social presence theory on online learning. Online Education and Adult Learning: New Frontiers for Teaching Practices, 124-139. Hershey, PA: IGI Global."
13. Ma, M., & Agarwal, R. (2007). Through a Glass Darkly : Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities Through a Glass Darkly : Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. Informatoin Systems Research, 18(1), 42–67.
14. Maher, B. (2016). Can a video game company tame toxic behaviour? Nature, 531(7596), 568–571.
15. Mennecke, B. E., Triplett, J. L., Hassall, L. M., & Heer, R. (2011). An Examination of a Theory of Embodied Social Presence in Virtual Worlds. Decision Sciences, 42(2), 413–450.
16. Riemer, H., & Shavitt, S. (2011). Impression management in survey responding: Easier for collectivists or individualists? Journal of Consumer Psychology, 21(2), 157–168.
17. Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London: John Wiley & Sons.
18. Suler, J. (2004). The Online Disinhibition Effect. CyberPsychology & Behavior, 7(3), 321–326.
19. Swann, W. B., J. T. Polzer, D. C. Seyle, & S. J. Ko. (2004). Finding value in diversity: Verification of personal and social self-view in diverse groups. Academy of Management Review. 29(1), 9-27.
20. Tu, C.-H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. The American Journal of Distance Education, 16(3), 131-150.
21. Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. Computers in Human Behavior, 41, 253–261.
22. Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. Computers in Human Behavior, 23(5), 2538–2557.
23. Yee, N. and Bailenson, J. (2007), The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. Human Communication Research, 33: 271–290.
24. Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. Computers in human behavior, 24(5), 1816-1836.