Journal of the Association for Information Systems

Volume 23 | Issue 3 Article 5

2022

When Socialization Goes Wrong: Understanding the We-Intention to Participate in Collective Trolling in Virtual Communities

Yang-Jun Li
City University of Hong Kong, yangjunli2-c@my.cityu.edu.hk

Christy M.K Cheung
Hong Kong Baptist University, ccheung@hkbu.edu.hk

Xiao-Liang Shen
Wuhan University, xlshen@whu.edu.cn

Matthew K. O. Lee
City University of Hong Kong, ismatlee@cityu.edu.hk

Recommended Citation
Li, Yang-Jun; Cheung, Christy M.K; Shen, Xiao-Liang; and Lee, Matthew K. O. (2022) "When Socialization Goes Wrong: Understanding the We-Intention to Participate in Collective Trolling in Virtual Communities," *Journal of the Association for Information Systems*, 23(3), 678-706.
DOI: 10.17705/1jais.00737
Available at: https://aisel.aisnet.org/jais/vol23/iss3/5

Journal of the Association for Information Systems (2022) **23**(3), 678-706
doi: 10.17705/1jais.00737

RESEARCH ARTICLE

ISSN 1536-9323

When Socialization Goes Wrong: Understanding the We-Intention to Participate in Collective Trolling in Virtual Communities

Yang-Jun Li,[1] Christy M. K. Cheung,[2] Xiao-Liang Shen,[3] Matthew K. O. Lee[4]

[1]College of Business, City University of Hong Kong, Hong Kong, China, yangjunli2-c@my.cityu.edu.hk
[2]School of Business, Hong Kong Baptist University, Hong Kong, China, ccheung@hkbu.edu.hk
[3]School of Information Management, Wuhan University, China, xlshen@whu.edu.cn
[4]College of Business, City University of Hong Kong, Hong Kong, China, ismatlee@cityu.edu.hk

Abstract

Although collective trolling poses a growing threat to both individuals and virtual community owners, the information systems (IS) literature lacks a rich theorization of this phenomenon. To address the research gaps, we introduce the concept of *we-intention* to capture the collective nature of collective trolling in virtual communities. We also integrate the social identity model of deindividuation effects (SIDE) and situational action theory to invoke the sociotechnical perspective in theorizing collective trolling in virtual communities. The objective of this study is to use the sociotechnical perspective to understand the we-intention to participate in collective trolling in virtual communities. We test our proposed model using data gathered from 377 Reddit users. Our moderated mediation analysis elaborates how technical elements (i.e., anonymity of self and anonymity of others) influence the we-intention to participate in collective trolling via individual-based social elements (i.e., perceived online disinhibition and social identity), with an environment-based social element (i.e., the absence of capable guardianship) as a boundary condition. We contribute to research by explaining collective trolling in virtual communities from the group-referent intentional action perspective and sociotechnical perspective. We also offer practical insights into ways to combat collective trolling in virtual communities.

Keywords: We-Intention, Collective Trolling, Social Identity Model of Deindividuation Effects (SIDE), Sociotechnical Perspective, Anonymity, Absence of Capable Guardianship, Collective Action, Moderated Mediation Analysis

Likoebe M. Maruping was the accepting senior editor. This research article was submitted on February 7, 2019 and underwent three revisions.

1 Introduction

Social technologies in the form of virtual communities (e.g., virtual worlds, social media, online support groups, and online discussion forums) connect geographically distant individuals and facilitate their pursuit of mutual interests and goals (Majchrzak & Malhotra, 2016; Karahanna et al., 2018; Mindel et al., 2018; Maruping et al., 2019). Individuals can obtain social and informational support by socializing and participating in public discussions in virtual communities. For example, people who experience health problems may join virtual health communities to obtain emotional support (Kordzadeh & Warren, 2017; Huang et al., 2019). In virtual brand communities, customers exchange product information and share shopping experiences with other customers and even brand owners (Yesiloglu et al., 2021). However, virtual communities are vulnerable to hostile social interactions such as online trolling,

which refers to a set of intentional, antisocial, and provocative online behaviors (Hardaker, 2010; Dineva & Breitsohl, 2021). Typical trolling behaviors in such communities include posting inflammatory, off-topic, or aggressive messages to provoke readers and disrupt online discussions (Sanfilippo et al., 2018).

The growing popularity of virtual communities has encouraged individual trollers to engage in collective trolling—whereby they coordinate trolling campaigns to harass other users and disrupt their online experiences (Al-khateeb & Agarwal, 2014, 2019; Krumsiek, 2017). On Reddit, for example, many trolling campaigns have been coordinated across subreddits of destructive and controversial communities[1] (Springer, 2015; Massanari, 2017). The subreddit /r/KotakuInAction, to which more than 96,000 members subscribe, is the main hub for the #GamerGate movement, which launches systematic trolling campaigns against female users (Massanari, 2017). This collective form of trolling, defined as "a form of collective action that involves an organized group trolling effort, targeting individuals or groups, while using trolling tactics and behaviors" (Sun & Fichman, 2020, p. 770), poses a serious threat to internet users, virtual community owners, and society. Collective trolling has more serious adverse consequences than individual trolling because victims experience multiple offenses from groups of people, making it difficult for them to fight back (Ransbotham et al., 2016). The negative consequences of collective trolling for victims include suicidal ideation, social anxiety, substance abuse, diminished life satisfaction, and delinquency (Citron, 2014, 2020; Krumsiek, 2017). These organized trolling behaviors also influence people's opinions on specific events, such as voting (Berghel & Berleant, 2018; Linvill et al., 2019; Zannettou et al., 2019). The insulting and manipulative posts used in collective trolling campaigns often provoke a flood of angry responses, resulting in a polarized society (Berghel & Berleant, 2018; Lew, 2019).

The emergence of collective trolling in virtual communities deserves more scholarly and public attention due to its devastating consequences for individuals and society. Nonetheless, our literature review identified research gaps that warrant further scholarly attention. First, prior studies have primarily conducted descriptive analyses of collective trolling (Flores-Saviaga et al., 2018; Kirkwood et al., 2019; Sun & Fichman, 2018, 2020; Ortiz, 2020); however, rich theorization of the phenomenon is lacking. This is a critical omission because collective trolling in virtual communities is a theoretically distinct phenomenon. Although its behavioral manifestations may resemble isolated individual deviant behaviors, collective trolling in virtual communities requires a different theorization. Specifically, since community members perceive themselves to be members of a larger social unit, their participation in collective trolling inevitably refers to other members and should thus be conceptualized as a group-referent intentional action (Bagozzi, 2000; Tsai & Bagozzi, 2014). Understanding the collective nature of this form of online deviant behavior requires investigation of the concept of *we-intention*, defined as "a commitment of an individual to participate in joint action [that] involves an implicit or explicit agreement between the participants to engage in that joint action" (Tuomela, 1995, p. 2).

Second, information systems (IS) researchers have not yet begun to investigate collective trolling in virtual communities. The field of IS, with its inherent focus on the sociotechnical view (Sarker et al., 2019), is predestined for exploring the role of social technologies in enabling collective trolling in virtual communities. The sociotechnical perspective considers both the social and technical elements of a phenomenon that is particularly suitable for addressing IT-related societal issues (Majchrzak et al., 2016). Most studies of collective trolling have treated technology as the research background without theorizing its impact on collective trolling (Kirkwood et al., 2019; Sun & Fichman, 2020; Ortiz, 2020). The technical design of a virtual community shapes members' psychological and motivational states (i.e., individual-based social elements), which may engage them in collective trolling in a virtual community. For example, enabling anonymity for members of virtual communities is a typical technical design decision made by virtual community owners to encourage users to express their views and generate content on the platforms (Scott & Orlikowski, 2014; Suh et al., 2018). However, an unintended consequence of this technical design choice is that it creates a safe psychological state (e.g., online disinhibition) for members to participate in collective trolling. In other words, collective trolling in virtual communities can result from technical design choices (i.e., anonymity) made by the virtual community owners.

Although anonymity has been shown to be a potent enabler of a wide range of antinormative and deviant behaviors (Postmes & Spears, 1998; Lowry et al., 2016), its role in collective trolling in virtual communities has not been theorized and tested. Furthermore, many virtual community owners use a wide range of countermeasures (e.g., moderators, policies, and reporting) to restrain users' deviant

[1] https://en.wikipedia.org/wiki/Controversial_Reddit_communities

behaviors. These countermeasures are widely considered to be environment-based social elements that intervene in users' decisions on deviant behaviors (Ransbotham & Mitra, 2009; Wikström, 2014). The novel application of the sociotechnical perspective to collective trolling can help IS researchers address the technical aspects (e.g., anonymity) of this phenomenon and consider its relationship with social elements (e.g., individual- and environment-based social elements). If we ignore this important theorization, we may generate an incomplete view of the causes of collective trolling and thus fail to provide effective technical and social interventions to address collective trolling in virtual communities.

Against these backdrops, this study theorizes collective trolling in virtual communities from the perspective of group-referent intentional action (Bagozzi & Dholakia, 2002, 2006a, 2006b; Tsai & Bagozzi, 2014) and the sociotechnical perspective (Sarker et al., 2019). Specifically, we seek to answer the following research questions:

RQ1: How does the perspective of group-referent intentional action contribute to our theoretical understanding of collective trolling in virtual communities?

To address this research question, we introduce the concept of we-intention to examine collective trolling participation in virtual communities. This concept captures the collective nature of this form of online deviant behavior.

RQ2: How does the sociotechnical perspective help to explain collective trolling in virtual communities?

We use the social identity model of deindividuation effects (SIDE) model (Reicher et al., 1995), which considers the distinctive role of anonymity in group behaviors, to theorize how technical elements trigger individual-based social elements in collective trolling in virtual communities. We also integrate situational action theory (Wikström, 2004, 2014) with the SIDE model to capture a salient favorable environment-based social element of collective trolling—the absence of capable guardianship (Chan et al., 2019; Wang et al., 2019)—that may intervene in users' decisions on whether to engage in collective trolling (Ransbotham & Mitra, 2009). By integrating the SIDE model and situational action theory, this study invokes the sociotechnical perspective for explaining the we-intention to engage in collective trolling in virtual communities. Furthermore, prior research has demonstrated the important role of person-environment interactions in crime and other deviant behaviors (Wikström, 2004, 2014). While the SIDE model captures technical elements (i.e., anonymity) and individual-based social elements (i.e., perceived online disinhibition and social identity), situational action theory serves as a complementary theory that highlights the boundary condition role of environment-based social elements (e.g., the absence of capable guardianship) in explaining users' participation in collective trolling in virtual communities (Ransbotham & Mitra, 2009; Wikström, 2014).

By theorizing the we-intention to participate in collective trolling in virtual communities from the group-referent intentional action perspective and sociotechnical perspective, we contribute to IS research and offer insights for owners of virtual communities. First, although IS researchers have begun to analyze deviant behaviors enabled by social technologies (e.g., Lowry et al., 2016, 2019; Chan et al., 2019; Wong et al., 2021; Wang & Lee, 2020), only minimal efforts have been made to explore collective deviant behaviors, such as collective trolling in virtual communities (Kirkwood et al., 2019; Sun & Fichman, 2020). Given the increasingly serious problems associated with collective trolling in virtual communities, this study advances the IS literature by providing a rich theorization of participation in collective trolling among members of virtual communities. The we-intention construct is highly relevant to IS research today because of the emergence of high-connectivity social technologies. Second, while governments and virtual community owners have begun to combat collective trolling, IS researchers have yet to join the conversation. We adopt the sociotechnical perspective (Sarker et al., 2019), which not only enables IS researchers to comprehensively engage with reference disciplines and address emerging societal issues but also informs virtual community owners in the ways that technical design and social elements can be used to combat collective trolling in their virtual communities. Thus, the results of this study offer insights into the construction and maintenance of a safe and healthy online environment for virtual community users.

2 Theoretical Foundation

This section first describes trolling and collective trolling and demonstrates how, from a sociotechnical perspective, collective trolling in virtual communities can be understood as a group-referent intentional action. We then introduce the SIDE model and situational action theory and explain how they invoke the sociotechnical perspective for theorizing collective trolling in virtual communities.

2.1 Trolling and Collective Trolling

Trolling is a catch-all term that comprises a set of intentional, antisocial, and provocative online behaviors intended to antagonize and upset others (Hardaker, 2010; Masui, 2019). Although trolling shares some similarities with cyberbullying (i.e., another widely

examined online deviant behavior), the two differ in several respects (Hardaker, 2010; Cruz et al., 2018). First, cyberbullying, defined as deliberate and aggressive behaviors intended to directly harm the victim (Lowry et al., 2016), is more direct and targeted than trolling; the victims of trolling are not necessarily predefined (Hardaker, 2010; Cruz et al., 2018). Second, cyberbullies tend to harass victims who cannot easily defend themselves (Chan et al., 2021). In contrast, trollers often provoke celebrities or politicians, who are likely to have more power than they do (Springer, 2015; Massanari, 2017). Finally, trolling is more purposive and strategic than cyberbullying, and often intends to provoke others and disrupt online interactions and public discussions (Coles & West, 2016; Masui, 2019). Thus, trolling has been observed in coordinated campaigns related to business and politics (Flores-Saviaga et al., 2018; Cruz et al., 2018).

There have been various recent trolling campaigns involving coordinated provocative and aggressive actions against certain individuals and groups in virtual communities (Springer, 2015; Massanari, 2017; Flores-Saviaga et al., 2018). This collective form of trolling has attracted increasing attention from the scientific community (Flores-Saviaga et al., 2018; Kirkwood et al., 2019; Sun & Fichman, 2018, 2020; Ortiz, 2020). Flores-Saviaga et al. (2018) examined the strategic use of collective trolling for political purposes and delineated the behavior patterns of groups of active trollers in a political trolling community on Reddit. Sun and Fichman (2020) conducted thematic content analysis of posts from a collective trolling event in China to characterize the lifecycle of collective trolling. While most studies in this area have focused on describing collective trolling, we provide a theoretical explanation of this phenomenon with a focus on group-referent intentional action (i.e., we-intention). We also fill a gap in IS research, which has not yet addressed collective trolling in virtual communities. Furthermore, we take a novel approach to collective trolling in virtual communities by adopting the sociotechnical perspective.

2.2 Approaching Collective Trolling in Virtual Communities from the Perspective of Group-Referent Intentional Action

Collective trolling is a set of coordinated trolling behaviors conducted by a group of people to provoke and upset other individuals or groups (Sun & Fichman, 2018, 2020). As a form of collective action, collective trolling requires a shared intention among group members that can be purely subjective and need not be true (Tuomela, 1995; Bagozzi, 2000). When individuals participate in collective trolling, they feel as if they are part of a group and that the collective trolling is conducted not by the individual alone but by a group of people. Accordingly, individuals' decisions to participate in collective trolling necessarily go beyond themselves to consider others in the group, particularly in virtual communities in which groups of like-minded people enjoy the experience of congregating and communicating as a group (Bagozzi & Dholakia, 2002). Some researchers have argued that action in relation to group members should be conceptualized differently from isolated individual actions, using such concepts as "shared intention," "collective intention," and "we-intention" (Gilbert, 1989; Bratman, 1992; Tuomela, 1995; Bagozzi, 2007). Bagozzi et al. conceptualized this kind of action as a group-referent intentional action and used the concept of we-intention to empirically examine virtual community members' intention to participate in a virtual community (Bagozzi & Dholakia, 2002, 2006a, 2006b; Bagozzi, 2000, 2007; Tsai & Bagozzi, 2014). Drawing on Bagozzi's series of studies, we use the concept of *the we-intention to participate in collective trolling*, defined as an individual community member's commitment to participate in collective trolling with the subjective perception that other members will also engage in the collective trolling (Tuomela, 1995), to theorize collective trolling in virtual communities.

The concept of we-intention has been increasingly applied to investigate the group-referent intentional use of social technologies such as virtual communities, groupware, social networking services, wikis, and multiplayer online games (see Table A1 in Appendix). These studies have suggested that the use of social technologies has meaning only when a group of users uses them in a collective manner (Cheung & Lee, 2010; Shen et al., 2010, 2011, 2021). Some studies have demonstrated that the formation of we-intention can facilitate collective actions among community members (Bagozzi & Dholakia, 2006a, 2006b), thus sustaining the community by alleviating concerns about free-riding (Bagozzi & Dholakia, 2006b; Mindel et al., 2018; Li & Suh, 2021). Furthermore, we find that most related studies have primarily used sociopsychological theories, such as social influence theory, the theory of planned behavior, and the model of goal-directed behavior (see Table A1 in Appendix), to explain we-intention. These sociopsychological theories have normally been used to examine the effects of social elements (e.g., individuals' psychological and motivational states and social environments) on we-intention, with technical elements serving as the hidden context (Sarker et al., 2019). However, we lack a rich understanding of the role of virtual communities in facilitating the we-intention to participate in collective trolling in virtual communities.

2.3 Approaching Collective Trolling in Virtual Communities from the Sociotechnical Perspective

Virtual communities are "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace" (Rheingold, 1993, p. 5). Similar to real-life communities, most virtual communities create and use a shared language, strive to achieve mutual goals, enact rituals, and establish boundaries that separate nonmembers via social technologies (Bagozzi & Dholakia, 2002). As such, virtual communities can be analyzed from the sociotechnical perspective, which describes "an ensemble, a practice, or even an analysis of any of these that integrates social and technical elements in a way that reveals their interactions and interpenetration" (Kling & Courtright, 2003, p. 222). Specifically, these social elements include individual-based social elements (e.g., individuals' attributes and psychological and motivational states) and environment-based social elements (e.g., the attributes of the environment in which individuals are situated). The technical elements include hardware and software and their associated resources and attributes (Sarker et al., 2019). The sociotechnical perspective is regarded as one of the foundational viewpoints in IS research (Bostrom & Heinen, 1977; Sarker et al., 2019) and it has been extensively used to theorize behaviors enabled by information technology (Lowry et al., 2016; Vaast et al., 2017; Karahanna et al., 2018; Chan et al., 2019; Wong et al., 2021).

We adopt the sociotechnical perspective to comprehensively consider key technical elements (anonymity of self and anonymity of others) and social elements (perceived online disinhibition, social identity, and the absence of capable guardianship) of collective trolling in virtual communities. The technical elements of virtual communities shape their individual-based social elements, which in turn provide a basis for collective trolling (Flores-Saviaga et al., 2018). Specifically, virtual communities connect groups of like-minded members with a keen awareness of their mutual interests and shared identity. Because of the anonymity offered by most virtual communities (an aspect of their technical design) (Gutman, 2018; Lowry et al., 2016), users are easily mobilized to participate in collective trolling against other individuals and groups via sociopsychological mechanisms (Flores-Saviaga et al., 2018). For instance, anonymity in virtual communities promotes the diffusion of responsibility because of a lack of identifiability and creates social pressure to conform to group norms (Postmes & Spears, 1998; Lowry et al., 2016; Krumsiek, 2017), thus facilitating users' engagement in collective trolling. In addition, virtual community owners should consider not only individual- but also environment-based social elements (e.g., guardianship) that may facilitate collective trolling. According to studies of crime and other deviant behaviors (Felson & Clarke, 1998; Wikström, 2004; Chan et al., 2019), potential offenders are more likely to exhibit deviant behaviors in the absence of capable guardianship. Most virtual communities use a variety of online countermeasures (e.g., peer-monitoring systems, moderators, policies and rules, and detection systems) to combat collective trolling. These countermeasures represent guardianship that deters collective trolling, and their effectiveness determines whether criminogenic opportunities for collective trolling arise in virtual communities.

2.4 Social Identity Model of Deindividuation Effects

The SIDE model is a useful framework for exploring the effects of social technologies on human in-group behaviors (Spears & Postmes, 2015; Spears, 2017). To capture the online representation of individuals and groups, we identify two distinctive roles of anonymity in the context of the virtual community: the anonymity of self and the anonymity of others.

The anonymity of self is defined as the extent to which individuals perceive themselves to be unidentifiable in online social interactions (Jiang et al., 2013). When people believe that they are unidentifiable and invisible to others (i.e., anonymity of self) in a virtual community, they can separate their online behaviors from their in-person identities and real personal lives (Suler, 2004). This provides them with several strategic advantages and further changes their behavioral decisions (Spears, 2017). For example, when communication technologies render individuals unidentifiable and invisible to others, they are likely to feel more relaxed and less restrained and to express themselves more openly because they are in a state of perceived online disinhibition (Cheung et al., 2021). Perceived online disinhibition refers to the sense of a lack of restraint when communicating with others online (Suler, 2004). Some individuals, however, use this sense of online disinhibition to engage in behaviors that violate socially accepted norms (e.g., cyberbullying, fraud, and trolling) (Postmes & Spears, 1998; Lowry et al., 2016; Huang et al., 2017; Kordyaka et al., 2020). In this study, the strategic mechanism of the SIDE model describes how the features of virtual communities create strategic advantages for members of a virtual community to participate in collective trolling, and perceived online disinhibition may be one of the major strategic advantages. The strategic mechanism assumes that members of a virtual community behave in certain ways that might be sensitive to power or socially accepted norms (Spears & Postmes, 2015; Spears, 2017).

The anonymity of others is defined as the extent to which an individual perceives others to be unidentifiable and anonymous in online social interactions (Jiang et al., 2013). When an individual has no personal information on other contributors to online group interactions, that individual lacks awareness of individual differences between the group members, making it difficult to distinguish between them (Lea et al., 2001; Lee, 2004). As a result, the individual views each group member not as a unique individual but as an interchangeable group member, amplifying the salience of social identity (Spears & Postmes, 2015; Spears, 2017). Social identity, defined as the self-awareness of the membership of a group and the emotional and evaluative significance of the membership (Tajfel, 1978), plays a key role in members' participation in collective actions in virtual communities (Ren et al., 2012; Laato et al., 2021). In this study, the cognitive mechanism of the SIDE model suggests that a lack of individuating information on other users obscures individual differences and leads to less individuated impressions, thus increasing group members' cognitive efforts to categorize themselves in terms of the group (Reicher et al., 1995; Spears, 2017).

2.5 Situational Action Theory

Situational action theory integrates environment-oriented and individual-oriented explanations of deviant behavior and thus emphasizes the contribution of person-environment interaction to such behaviors (Wikström, 2004, 2014). This theory states that deviant behaviors depend on the interaction between individual factors and criminogenic settings (Wikström, 2004; Pauwels et al., 2018). Individual factors that can explain offenders' tendencies to engage in deviant behavior include personal traits, capabilities, and psychological or motivational states (Wikström, 2014). Individuals with a lack of self-control, a dark tetrad personality, or perceived online disinhibition are likely to exhibit deviant behaviors online (Song & Lee, 2020; Masui, 2019; Wong et al., 2018; Kordyaka et al., 2020). Criminogenic settings are environments whose conditions are conducive to deviant behaviors—for example, guardianship conditions (e.g., a lack of supervision or deterrence measures) (Wikström, 2014; Pauwels et al., 2018) that encourage offenders to engage in deviant behaviors (Ransbotham & Mitra, 2009; D'Arcy et al., 2009). Offenders are likely to rationally avoid deviant behavior when the risk of being caught and punished is high, even when they have a strong motivation to engage in such behaviors (see Pauwels et al., 2018 for a review). Based on situational action theory, which emphasizes the role of person-environment interaction in explaining deviance, environment-based social elements, such as the absence of capable guardianship, are likely to interact with individual-based social elements, such as perceived online disinhibition and social identity, and lead to the we-intention to participate in collective trolling in virtual communities.

3 Research Model and Hypothesis Development

Figure 1 depicts the research model of the we-intention to participate in collective trolling in virtual communities. Guided by the group-referent intentional action literature and the sociotechnical perspective, we introduce the concept of we-intention and use the SIDE model and situational action theory to invoke a sociotechnical perspective for theorizing collective trolling in virtual communities. Specifically, key technical elements (i.e., anonymity of self and anonymity of others) trigger key individual-based social elements (i.e., perceived online disinhibition and social identity) via the strategic and cognitive mechanisms of the SIDE model. Based on situational action theory, these individual-based social elements interact with an environment-based social element (i.e., the absence of capable guardianship) to determine the we-intention to participate in collective trolling in virtual communities. Moreover, prior research suggests that online trolling behaviors vary across demographic groups (e.g., age, gender, education, and community experience) (Craker & March, 2016; Ferguson & Glasgow, 2021). Following prior research on trolling (Masui, 2009) and other online deviant behaviors (e.g., cyberbullying) (Lowry et al., 2016; Chan et al., 2019), we also include several demographic characteristics (e.g., age, gender, education, and community experience) as control variables to ensure the robustness of the research model.

3.1 Strategic Mechanism of SIDE Model: The Mediating Role of Perceived Online Disinhibition

The strategic mechanism of the SIDE model describes how the advantages provided by communication technologies (e.g., anonymity of self) influence individuals' behaviors (Reicher et al., 1995; Spears & Lea, 1994). Through the affordances of communication technologies, most virtual communities (e.g., Reddit) allow users to manipulate their identifiability and visibility by managing the displays of their digital profiles and their privacy settings. When members of virtual communities are given the option of anonymity, they are likely to loosen up and express themselves more openly (i.e., the online disinhibition effect) because the anonymity of self allows them to separate their online behaviors from their personal identity and gives them the sense that they cannot be easily identified (Suler, 2004; Huang et al., 2017).

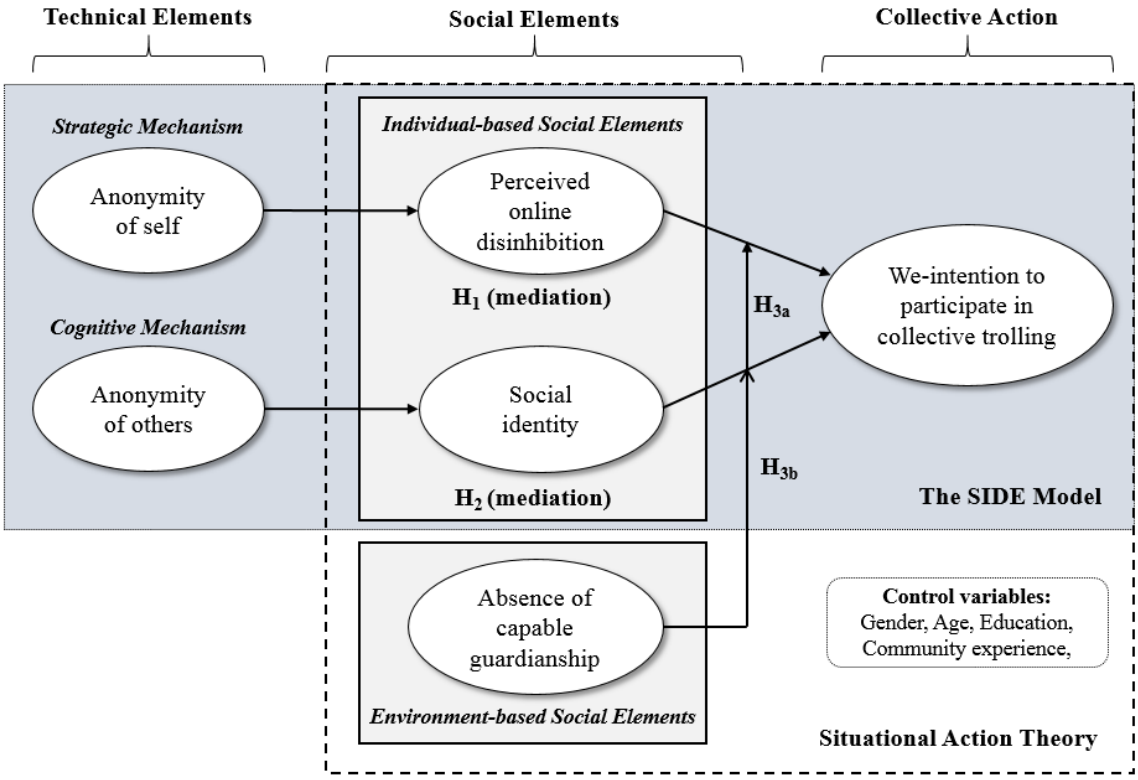


Figure 1. A Sociotechnical Model of Collective Trolling in Virtual Communities

Perceived online disinhibition has been regarded as one of the most direct results of the anonymity of self (Suler, 2004; Lowry et al., 2016; Cheung et al., 2021). Furthermore, social interactions and group decisions can easily become extreme and irrational because of the diffusion of responsibility entailed by group decision-making (Sia et al., 2002; Laato et al., 2021), especially for groups comprising largely like-minded people who are easily mobilized in the absence of restraints to combine their efforts to provoke others (Krumsiek, 2017; Strandberg et al., 2019). Previous studies largely agree that perceived online disinhibition is accountable for a wide variety of online behaviors that violate socially accepted norms (Hardaker, 2010; Wong et al., 2018; Harrison, 2018; Kordyaka et al., 2020). When members of a virtual community feel disinhibited, they believe that their behaviors are not accountable to other members and they are likely to seize upon this strategic advantage by engaging in community interactions to collectively provoke and upset others.

Overall, when members of virtual communities are allowed to remain anonymous and unidentifiable to other members during their community interactions, they will experience reduced inhibition and restraint, which offers them a safe psychological state in which to actively interact with other members and form we-intentions to engage in collective trolling in virtual communities. Accordingly, in line with the strategic mechanism of the SIDE model (Reicher et al., 1995), perceived online disinhibition can be regarded as one of the strategic advantages provided by the anonymity of self, which further facilitates the formation of the we-intention to participate in collective trolling in virtual communities. That is, the formation of the we-intention to participate in collective trolling in a self-anonymous community may result in part from the perceived online disinhibition triggered by the anonymity of self. It is not the anonymity of self per se but the perceived online disinhibition caused by the anonymity of self that fosters the we-intention to participate in collective trolling. Therefore, perceived online disinhibition operates as a mediating mechanism between the anonymity of self and the we-intention to participate in collective trolling in virtual communities. Studies taking the sociotechnical perspective have also suggested that an individual's psychological state associated with the use of information technology plays a key role in the internalization of technological features to form behavioral intentions (Wixom & Todd, 2005; Sarker et al., 2019; Yang & Gong, 2021). For example, Ren et al. (2012) demonstrated that members' psychological attachment mediates the impact of virtual community features on members' participation in virtual communities. Therefore, we hypothesize:

H1: Perceived online disinhibition mediates the effect of the anonymity of self on the we-intention to participate in collective trolling in virtual communities.

3.2 Cognitive Mechanism of the SIDE Model: The Mediating Role of Social Identity

The cognitive mechanism of the SIDE model describes the cognitive processes by which the absence of individuating information (i.e., anonymity of others) in group interactions influences the salience of social identity and in-group behaviors (Reicher et al., 1995; Spears & Lea, 1994). In our study, social identity refers to a virtual community member's awareness of their membership as a community member and the emotional and evaluative significance assigned to this membership (Tajfel, 1978). In most virtual communities, members can obtain other members' personal information by checking their profiles (e.g., email, username, and birthday) and historical posting and/or chatting behaviors (Jiang et al., 2013). Ma and Agarwal (2007) identified four virtual community artifacts (i.e., virtual copresence, persistent labeling, self-representation, and deep profiling) that can influence community members to obtain individuating identity information about other members. According to the cognitive mechanism of the SIDE model, when virtual communities prevent community members from obtaining individuating information about others, members cannot determine how others differ from themselves; as a result, they will see other members as interchangeable rather than as unique individuals (Spears & Postmes, 2015; Spears, 2017). In this case, members' cognitive efforts to perceive the virtual community as an entity will be amplified, increasing their tendency to categorize themselves in terms of the community (i.e., social identity with the virtual community) (Lea et al., 2001; Lee, 2004).

When community members develop a sense of social identity with the community, they evaluate the community in a positive manner and exert efforts to participate in social interactions with others to maintain their membership (Tajfel, 1978; Ren et al., 2012), providing the core conditions for the formation of a we-intention to join collective actions (Bagozzi & Dholakia, 2002). In particular, groupthink—a psychological phenomenon occurring within a group of people who have a desire for harmony or conformity—may arise, encouraging members to stereotype anyone outside the group as an enemy (Turner & Pratkanis, 1998; Krumsiek, 2017). Those involved in groupthink may also temporarily forget their own ideas and conform to group decisions in social interactions (Krumsiek, 2017; Spears & Postmes, 2015). As a result, community members with a strong sense of social identity are likely to follow other members and form groups to provoke and attack others, thus forming a strong we-intention to participate in collective trolling in virtual communities. Studies have identified social identity as a key antecedent of the we-intention to participate in collective actions in the community (Bagozzi & Dholakia, 2002, 2006a; Cheung & Lee, 2010).

Accordingly, in line with the cognitive mechanism of the SIDE model (Reicher et al., 1995), the anonymity of others afforded by virtual communities obscures individual differences between community members and increases the salience of social identity with the community. This facilitates the formation of the we-intention to participate in collective trolling to maintain a sense of social identity in the virtual community. It is plausible that, in the absence of individuating information about other members, people will form we-intentions to participate in collective trolling only when they feel that they identify with the community. That is, social identity triggered by the anonymity of others, rather than the anonymity of others per se, facilitates the formation of the we-intention to participate in collective trolling in virtual communities. Therefore, social identity plays a key mediating role in transforming the anonymity of others into the we-intention to participate in collective trolling in virtual communities. This is consistent with the sociotechnical perspective that suggests that technologies influence human behaviors by causing psychological and motivational changes in individuals (Wixom & Todd, 2005; Sarker et al., 2019; Yang & Gong, 2021). Studies have also suggested that identification with a virtual community mediates the effects of the community's features on participation behaviors (Ren et al., 2012). Therefore, we hypothesize:

H2: Social identity mediates the effect of the anonymity of others on the we-intention to participate in collective trolling in virtual communities.

3.3 Moderating Role of the Absence of Capable Guardianship

According to situational action theory, engagement in deviant behaviors depends on two factors: a motivated offender who is ready or willing to commit deviant behaviors (i.e., individual-based social elements) and an environment that offers opportunities for such behaviors (i.e., environment-based social elements) (Wikström, 2004, 2014). Every deviant behavior requires an opportunity but not every opportunity will result in deviant behavior (Wikström, 2004, 2014). Moreover, individual-based social elements that help explain deviant behaviors (e.g., psychological and motivational states) are necessary but not sufficient conditions (Felson & Clarke, 1998). For example, if environmental conditions are not conducive to deviant behaviors, then potential offenders may rationally refrain from engaging in deviant behaviors to avoid being caught or punished, even if they are motivated or prepared to engage in such behaviors (Felson & Clarke, 1998; Chan et al., 2019). As such, environmental conditions limit the effects of individual factors on the decision to engage in deviant behavior.

When no capable guardianship (e.g., monitoring and deterrence measures) is present to combat collective trolling in a virtual community, the guardianship environment is amenable to collective trolling because of the low risk of being punished (D'Arcy et al., 2009; Pauwels et al., 2018; Chan et al., 2019). Virtual community members are likely to seize on this criminogenic opportunity to translate the strategic advantages provided by the anonymity of self (i.e., perceived online disinhibition) into collective trolling behaviors, resulting in a stronger we-intention to participate in collective trolling in virtual communities (Felson & Clarke, 1998; Chan et al., 2019). That is, in the absence of capable guardianship, disinhibited members will form greater we-intentions to participate in collective trolling in virtual communities. This indicates a stronger mediating role of perceived online disinhibition between the anonymity of self and the we-intention to participate in collective trolling. In contrast, when members feel that in-community guardianship to deter collective trolling is capable, they will expect punishment if they engage in collective trolling. Thus, members tend to be deterred by capable guardianship, which leads them to rationally relinquish the strategic advantages provided by the anonymity of self (i.e., perceived online disinhibition) and reduce their participation in collective trolling. This decreases the we-intention to participate in collective trolling in virtual communities. In this case, the mediating role of perceived online disinhibition between the anonymity of self and the we-intention to participate in collective trolling is weakened. This line of reasoning suggests that the indirect effect of the anonymity of self on the we-intention to participate in collective trolling via perceived online disinhibition depends on the effectiveness of guardianship as a deterrent to collective trolling. Therefore, we hypothesize:

H3a: The absence of capable guardianship moderates the mediating effect of perceived online disinhibition between the anonymity of self and the we-intention to participate in collective trolling such that the mediating effect will be stronger when the absence of capable guardianship is high compared to when it is low.

We further hypothesize that the absence of capable guardianship moderates the mediating effect of social identity on the relationship between the anonymity of others and the we-intention to participate in collective trolling in virtual communities. When members of a virtual community feel that the community offers no capable guardianship to deter collective trolling, they likely will not expect their participation in collective trolling to be punished. This offers a criminogenic opportunity for members to engage in collective trolling in virtual communities (D'Arcy et al., 2009; Pauwels et al., 2018; Chan et al., 2019). In this case, members who identify with the community will be more willing to convert their collective trolling motivations derived from the anonymity of others (i.e., social identity) into collective trolling behaviors, without the fear of being monitored and punished (Felson & Clarke, 1998; Chan et al., 2019). That is, in the absence of capable guardianship, social identity exerts a stronger effect on the we-intention to participate in collective trolling, yielding a stronger indirect effect of the anonymity of others on the we-intention to participate in collective trolling via social identity. In contrast, when the virtual community has capable guardianship to deter collective trolling, members who identify with the community may reduce their participation in collective trolling in virtual communities to reduce the risk of being punished. In this case, the indirect effect of the anonymity of others on the we-intention to participate in collective trolling is reduced because members are less likely to translate their motivation (i.e., social identity) into collective trolling behaviors (i.e., the we-intention to participate in collective trolling). Based on these arguments, we expect that the indirect effect of the anonymity of others on the we-intention to participate in collective trolling via social identity depends on the degree of the absence of capable guardianship in the context of collective trolling. Therefore, we hypothesize:

H3b: The absence of capable guardianship moderates the mediating effect of social identity between the anonymity of others and the we-intention to participate in collective trolling such that the mediating effect will be stronger when the absence of capable guardianship is high compared to when it is low.

4 Research Methods

This section introduces the research setting, measures, data collection procedure, profiles of the respondents, and potential response biases.

4.1 Research Setting

To test our research model, we recruited Reddit users to participate in an online survey. Online surveys have been extensively used to examine various online deviant behaviors, such as cyberbullying, cyberharassment (Lowry et al., 2016; Chan et al., 2019), and online fraud behaviors (Harrison, 2018). The target population group for our survey comprised users who (1) were subreddit users and (2) had joined certain subreddits on which collective trolling had been common during the past 6 months. Reddit is a well-known virtual community website that comprises more than 2 million user-created subcommunities of interest (i.e., subreddits) and has approximately 430 million active monthly users (as of July 2020). Unlike

Twitter and Facebook, Reddit favors aliases over real names and is committed to the defense of internet anonymity (Gutman, 2018). It is notorious for hosting subreddits on which trollers share and upvote content that is offensive to others, resulting in collective trolling (Springer, 2015; Massanari, 2017). Reddit has also been selected as the research setting for several other studies of online collective deviant behaviors (Massanari, 2017; Flores-Saviaga et al., 2018). Therefore, we considered Reddit to be an appropriate context in which to test our model.

4.2 Measures

We used previously validated items to measure major constructs in the model. We measured the anonymity of self and the anonymity of others using items adapted from Lowry et al. (2016). We adapted items from Wong et al. (2018) to measure perceived online disinhibition. Social identity was measured using items adapted from Ren et al. (2012). We measured the degree of capable guardianship using items adapted from Chan et al. (2019). We adapted items from Tsai and Bagozzi (2014) and Shen et al. (2014) to measure the we-intention to participate in collective trolling. We modified these items to fit the context of collective trolling where necessary. Table 1 presents the measurement items. We operationalized all of the constructs as reflective constructs because all of the items assigned to the same construct were highly correlated (Jarvis et al., 2003; MacKenzie et al., 2011). We used a 7-point Likert "disagree-agree" scale to frame all of the items. Given our interest in socially undesirable behavior (collective trolling), we included the social desirability scale proposed by Reynolds (1982) in the questionnaire to detect potential response bias. Before distributing the questionnaire, we pretested the designed instruments with several IS researchers and Reddit users to ensure the items' face validity. We asked the pretest subjects to complete the entire survey and to comment on the items' simplicity, precision, and clarity. We used the subjects' feedback to revise problematic items in the final questionnaire.

4.3 Data Collection

We recruited our Reddit sample from Amazon Mechanical Turk (MTurk). MTurk is an online crowdsourcing platform on which registered users with diverse characteristics (e.g., gender, age, education, job, and income) can participate in compensated survey tasks. It thus allowed us to obtain a highly representative sample of Reddit users. The use of MTurk, a globally renowned professional third-party portal website, also tends to reduce respondents' concerns regarding privacy, which facilitates t eliciting honest responses related to sensitive research topics (here, collective trolling). We followed the methodological guidelines on MTurk to design and distribute our survey (Steelman et al., 2014) and designed several screening questions to ensure that the respondents were Reddit users that had joined subreddits on which collective trolling had been prevalent during the previous six months. For example, we asked the respondents to select up to three social media platforms they had used most frequently during the past six months. Only Reddit users were allowed to continue the survey. We further asked the respondents whether they had seen collective trolling activities on their frequently visited subreddits and asked them to provide some basic information about that subreddit (e.g., name, URL, and the number of subscribers). We randomly searched for some of these subreddits to verify that they existed. We also included several attention-check questions in the survey to ensure response quality: four questions asked the respondents to choose the specified option, and one question asked the respondents to make a choice based on their gender.

4.4 Respondent Profiles

After filtering the data on the basis of the screening questions and attention-check questions,[2] we had 377 valid responses. Table 2 details the respondents' profiles. Approximately half of the respondents (52.79%) were between 25 and 34 years old, 66.84% were male, and 64.99% and 25.99% held bachelor's and master's degrees, respectively. Most of the respondents had full-time jobs (90.45%). Most had contributed to the focal subreddit for more than six months (85.68%). Most visited the focal subreddit frequently, with 22.02% visiting more than once a day and 30.24% visiting once per day. The respondents' characteristics were consistent with the overall profile of Reddit users, most of whom are young men (Sattelberg, 2020).

4.5 Response Bias Detection

Common method bias (CMB) is a concern when collecting data from the same source at the same time through an online survey (Lindell & Whitney, 2001; Podsakoff et al., 2003). We assessed the risk of CMB in several ways. First, all of the correlations between the major constructs in the model fell significantly below 0.90 (see Table 3), suggesting that CMB was not present (Lowry et al., 2016).

[2] A total of 8,294 respondents visited the URL for our questionnaire, 141 of whom did not agree to take the survey. We asked the respondents to first answer all of the screening questions; this left 581 respondents who answered the questions related to the major constructs. The response rate was 7.13%. As 204 respondents failed the attention-check questions, our final sample contained 377 valid responses.

Table 1. Constructs and Measurement Items

Anonymity of self (AS), adapted from Lowry et al. (2016)
AS1: Reddit will not identify me without my permission. AS2: My personal identity information will NOT be attached to the internal records of Reddit unless that is what I want.
Anonymity of others (AO), adapted from Lowry et al. (2016)
AO1: Reddit will not identify other subreddit members and divulge their personal identities without their permission. AO2: No personal identity information will be attached to the internal records of Reddit unless other subreddit members permit.
Perceived online disinhibition (POD), adapted from Wong et al. (2018)
POD1: I feel less nervous when sharing personal ideas or feelings in the subreddit. POD2: I feel like I can be more open when I am communicating in the subreddit. POD3: I feel like I can sometimes be more personal during conversations in the subreddit. POD4: When online, I feel more comfortable disclosing personal information to a member of the opposite sex in the subreddit. POD5: I feel less shy when I am communicating in the subreddit. POD6: I feel less embarrassed sharing personal ideas or feelings with another person in the subreddit.
Social identity (SI), adapted from Ren et al. (2012)
SI1: I identify with the subreddit. SI2: I feel connected to the subreddit. SI3: I feel I am a typical member of the subreddit.
Absence of capable guardianship (ACG), adapted from Chan et al. (2019)
ACG1: The guardianships of moderators in the subreddit to prevent collective trolling are not capable. ACG2: The guardianships of moderators in the subreddit to deter collective trolling are ineffective. ACG3: The guardianships of moderators in the subreddit to regulate collective trolling are not competent. ACG4: The guardianships of moderators in the subreddit to tackle collective trolling are ineffective.
We-intention to participate in collective trolling (WPCT), adapted from Tsai and Bagozzi (2014) and Shen et al. (2014)
WPCT1: I intend that our subreddit members troll someone together in the subreddit during the next few months. WPCT2: We [i.e., my subreddit members and I] intend to troll someone together in the subreddit during the next few months. WPCT3: I believe that we [i.e., my subreddit members and I] mutually agree to troll someone together in the subreddit during the next few months. WPCT4: I believe that we [i.e., my subreddit members and I] share a common intention to troll someone together in the subreddit during the next few months.

Table 2. Profiles of Respondents

Characteristics	***N***	**%**	**Characteristics**	***N***	**%**
Gender			Duration of each visit to the focal subreddit		
Male	252	66.84	Less than 5 minutes	14	3.71
Female	125	33.16	5-10 minutes	66	17.51
Age			11-20 minutes	136	36.07
18-24	54	14.32	21-30 minutes	118	31.30
25-34	199	52.79	30-60 minutes	26	6.90
35-44	54	14.32	More than 60 minutes	17	4.51
45-54	61	16.18	Frequency of visiting the focal subreddit		
55 or older	9	2.39	More than once every day	83	22.02
Education			Once per day	114	30.24
High school (equivalent)	7	1.86	A few times a week	135	35.81
Associate degree	6	1.59	A couple of times a month	29	7.69
Bachelor degree	245	64.99	A few times a year	11	2.92
University (no degree)	17	4.51	Rarely	5	1.33
Master or higher	102	27.05	Experience of joining the focal subreddit		
Employment status			Less than 6 months	54	14.32
Student	11	2.92	6-12 months	156	41.38
Full-time	341	90.45	1-2 years	124	32.89
Part-time	19	5.04	3-5 years	32	8.49
Unemployed	6	1.59	More than 5 years	11	2.92
Annual income (US Dollars)			Experience of using social media service		
Less than $10,000	36	9.55	Less than 6 months	25	6.63
$10,001 to $30,000	67	17.77	6-12 months	53	14.06
$30,001 to $50,000	98	25.99	1-2 years	102	27.06
$50,001 to $70,000	79	20.95	3-5 years	129	34.22
$70,001 to $100,000	69	18.30	6-10 years	40	10.61
More than $100,000	28	7.43	More than 10 years	28	7.43

Second, we performed Harman's single-factor test, which has been widely used for CMB testing (Podsakoff et al., 2003; Srivastava & Chandra, 2018). The results of the principal component analysis suggested that the extracted primary component accounted for less than half of the variance in the data (40.46%), indicating the CMB was not a threat (Podsakoff et al., 2003). Third, we adopted the marker variable technique to control for potential CMB, as recommended by Lindell and Whitney (2001). This technique can address most of the problems related to Harman's single-factor test (Lindell & Whitney, 2001; Podsakoff et al., 2003). For example, a single factor is less likely to explain the majority of the variance in the data if the number of latent variables in the model is higher (Srivastava & Chandra, 2018). The marker variable technique requires researchers to include a variable that theoretically bears no relation to the constructs in the model (Lindell & Whitney, 2001). Alternatively, researchers can estimate CMB in a post hoc fashion without the need to carefully identify a marker variable before data collection (Lindell & Whitney, 2001; Richardson et al., 2009). The second-smallest positive correlation among the manifest variables is regarded as a reasonable and conservative estimate of CMB (Lindell & Whitney, 2001; Malhotra et al., 2006). Many IS researchers have estimated CMB in a post hoc fashion (Durcikova et al., 2018; Liang et al., 2019; Shen et al., 2019). Because we did not include a theoretically unrelated construct (i.e., marker variable) in the model when we administered our survey, we used the post hoc marker variable technique. As shown in Table 3, the second-smallest positive correlation among the manifest variables was 0.011. We then followed the procedure detailed by Lindell and Whitney (2001) to partial out the CMB estimate (i.e., 0.011) from the previously observed correlations for which contamination by CMB was suspected. The results indicated that the significance levels of the correlations among the variables remained nearly unchanged. Therefore, we concluded that CMB was not a serious threat in our study.

The use of self-reported responses to examine deviant behaviors—collective trolling in our study—may produce social desirability bias (SDB) (Kwak et al., 2019). We applied several measures to minimize this threat. First, in order to alleviate their concerns and encourage them to respond honestly, the participants were informed that the survey was entirely anonymous and voluntary. Second, as suggested by Reynolds (1982), we collected data on social desirability from the survey respondents. If Spearman's correlation between the SDB score and the dependent variable of interest is significant and negative, SDB represents a threat (Chan et al., 2019). Our results show that the Spearman rho value between SDB and the we-intention to participate in collective trolling was −0.060 ($p > 0.05$), which is not significant. Therefore, SDB is not a concern for our study.

5 Data Analysis and Results

We used the structural equation modeling (SEM) approach and partial least squares (PLS) technique to test the main model. PLS-SEM is more robust than covariance-based SEM (CB-SEM); it has fewer statistical identification issues and is more suitable for testing models with relatively small samples (Hair et al., 2011), as in our study. Studies have also indicated that PLS-SEM and CB-SEM yield similar parameter estimations (Hair et al., 2011). Moreover, PLS-SEM can mitigate concerns associated with regression analysis (e.g., measurement errors) when used to estimate highly complex mediation models with latent variables (such as the moderated mediation model in our study) (Sarstedt et al., 2020). Following a two-step analytical approach (Hair et al., 2011), we first examined the measurement model to assess the reliability and validity of the measures and then examined the structural model to test our hypotheses. To cross-validate the results of PLS-SEM, we further applied the PROCESS macro in SPSS to test the mediation and moderation hypotheses.

5.1 Measurement Model Evaluation

Following Hair et al. (2011), we assessed construct reliability, convergent validity, and discriminant validity to evaluate the psychometric properties of the measures in the reflective measurement models. Composite reliability measures the internal consistency of a construct to evaluate its reliability (Fornell & Larcker, 1981; Hair et al., 2011). A composite reliability value greater than 0.7 is considered satisfactory (Hair et al., 2011). The results in Table 4 indicate that the composite reliability values ranged from 0.836 to 0.900, demonstrating good construct reliability (Hair et al., 2011).

Convergent validity measures the degree to which construct items that should be theoretically related are actually related (Fornell & Larcker, 1981; Hair et al., 2011). We assessed convergent validity using two criteria: (1) average variance extracted (AVE) values should be greater than 0.5, indicating that the latent construct can explain more than half of the variance in its items; and (2) all item loadings on the theoretically assigned constructs should be greater than 0.7 (Fornell & Larcker, 1981; Hair et al., 2011). As shown in Table 4, the AVE values range from 0.580 to 0.753 and all item loadings on the given constructs are greater than 0.7 (bold numbers in Table 4), except POD5. Therefore, the constructs show satisfactory convergent validity. Discriminant validity measures the extent to which items that are not supposed to be theoretically related are indeed unrelated (Fornell & Larcker, 1981; Hair et al., 2011). We used two measures of discriminant validity: (1) the Fornell-Larcker criterion and (2) the cross-loading criterion (Fornell & Larcker, 1981; Hair et al., 2011).

Table 3. Correlation Matrix

	Mean	SD	VIF	Age	Edu	Gen	CE	ACG	AO	AS	POD	SI	WPCT
Age	3.387	1.020	1.096										
Edu	5.111	0.814	1.195	0.198***									
Gen	1.326	0.469	1.072	0.119*	0.141**								
CE	2.443	0.938	1.036	0.051	0.099	−0.039							
ACG	5.149	1.159	1.694	−0.088	0.198***	0.030	0.011	**0.832**					
AO	5.247	1.107	1.794	−0.056	−0.036	0.017	0.082	0.393***	**0.868**				
AS	5.206	1.067	1.909	−0.063	0.047	0.049	0.013	0.372***	0.596***	**0.847**			
POD	5.166	1.013	2.639	−0.042	0.118*	0.146**	0.074	0.582***	0.531***	0.575***	**0.762**		
SI	5.354	1.090	2.077	−0.124*	0.107*	0.080	0.074	0.407***	0.471***	0.514***	0.662***	**0.845**	
WPCT	5.283	1.023	1.797	0.000	0.242***	0.019	0.019	0.485***	0.412***	0.463***	0.542***	0.548***	**0.805**

Note: 1. Edu = education, Gen = gender, CE = community experience, ACG = absence of capable guardianship, AO = anonymity of others, AS = anonymity of self, POD = perceived online disinhibition, SI = social identity, WPCT = we-intention to participate in collective trolling, SD = standard deviation, VIF = variance inflation factor. 2. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. 3. The diagonal elements in bold represent the square root of AVE.

Table 4. Construct Reliability and Validity

Constructs	AVE	CR	Items	ACG	AO	AS	POD	SI	WPCT
Absence of capable guardianship (ACG)	0.693	0.900	ACG1	**0.846**	0.340	0.302	0.514	0.343	0.394
			ACG2	**0.821**	0.336	0.340	0.513	0.336	0.428
			ACG3	**0.813**	0.308	0.327	0.428	0.303	0.399
			ACG4	**0.850**	0.321	0.274	0.487	0.372	0.395
Anonymity of others (AO)	0.753	0.859	AO1	0.369	**0.861**	0.516	0.453	0.399	0.349
			AO2	0.314	**0.875**	0.525	0.471	0.420	0.367
Anonymity of self (AS)	0.718	0.836	AS1	0.300	0.445	**0.827**	0.452	0.392	0.380
			AS2	0.333	0.566	**0.867**	0.528	0.478	0.407
Perceived online disinhibition (POD)	0.580	0.892	POD1	0.428	0.438	0.428	**0.780**	0.527	0.382
			POD2	0.489	0.444	0.508	**0.797**	0.524	0.434
			POD3	0.363	0.376	0.447	**0.733**	0.475	0.373
			POD4	0.447	0.371	0.393	**0.760**	0.449	0.458
			POD5	0.439	0.345	0.381	**0.686**	0.506	0.385
			POD6	0.496	0.448	0.481	**0.807**	0.541	0.445
Social identity (SI)	0.714	0.882	SI1	0.357	0.381	0.487	0.591	**0.839**	0.478
			SI2	0.341	0.385	0.440	0.564	**0.855**	0.452
			SI3	0.333	0.429	0.382	0.521	**0.841**	0.459
We-intention to participate in collective trolling (WPCT)	0.648	0.881	WPCT1	0.360	0.302	0.371	0.463	0.428	**0.810**
			WPCT2	0.380	0.373	0.424	0.454	0.443	**0.801**
			WPCT3	0.420	0.318	0.371	0.429	0.458	**0.812**
			WPCT4	0.404	0.334	0.328	0.403	0.435	**0.797**

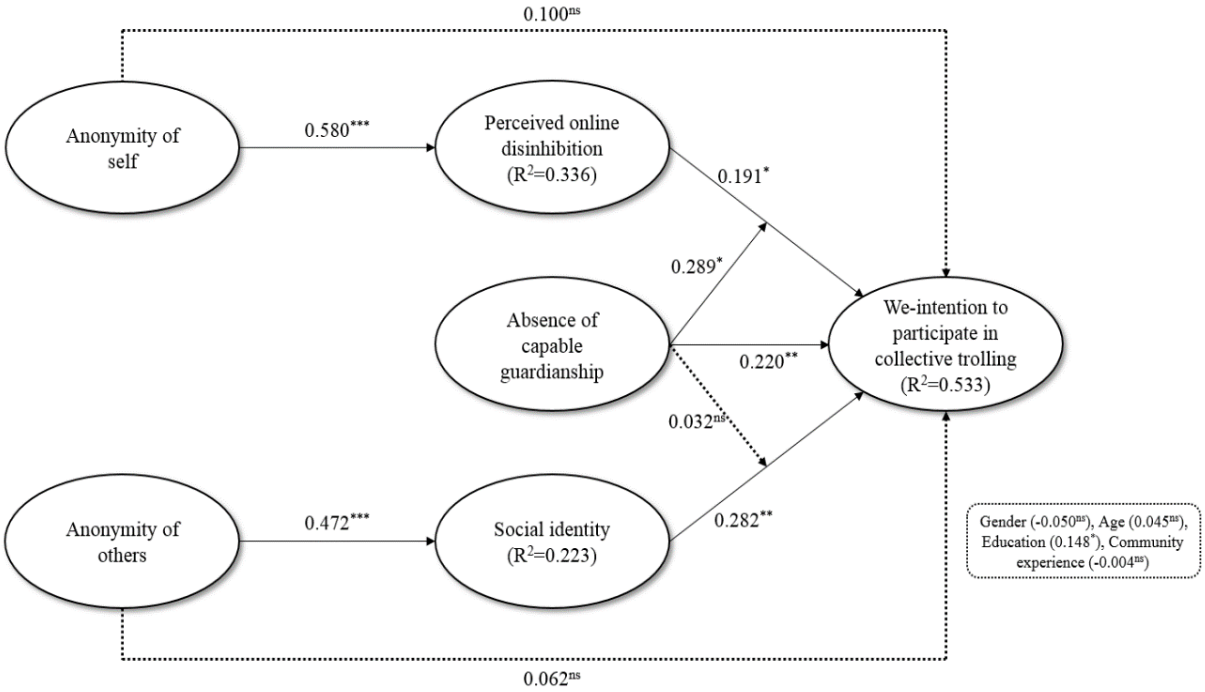
Note: AVE = average variance extracted, CR = composite reliability. The diagonal elements in bold represent the item loadings on the theoretically assigned constructs.

To meet the Fornell-Larcker criterion, a latent construct should share more variance with its theoretically assigned items than with any other latent construct in the model (Hair et al., 2011). Therefore, the square root of each AVE should be greater than the correlations between that AVE and all other constructs. To meet the cross-loading criterion, the items of each construct should load higher on the associated construct than on all remaining constructs (Hair et al., 2011). Tables 3 and 4 show that every construct had good discriminant validity.

We also calculated variance inflation factor values to assess potential problems with multicollinearity. Table 3 shows that all of the variance inflation factor values ranged from 1.036 to 2.639, falling below the stringent threshold of 3 (Diamantopoulos, 2011), demonstrating that our data were not seriously affected by multicollinearity.

5.2 Hypothesis Testing

We ran the structural model in SmartPLS 2.0 to estimate the path coefficients and associated *t*-values, using the bootstrapping method with 1,000 iterations. Figure 2 shows the PLS-SEM results.



Note: ******* $p < 0.001$; ****** $p < 0.01$; ***** $p < 0.05$; ns = not significant.

Figure 2. PLS-SEM Results

The main model explained 53.3% of the variance in the we-intention to participate in collective trolling, with 33.6% of the variance in perceived online disinhibition and 22.3% of the variance in social identity. We also evaluated the model fit by calculating the goodness-of-fit (GoF) value, which was recommended by Wetzels et al., (2009) and by Henseler and Sarstedt (2013) for assessing the predictive ability of PLS-SEM. The GoF value is defined as the geometric mean of the average communality of all major constructs and the average R^2 of each of the endogenous constructs in the model (Wetzels et al., 2009). In line with the effect sizes of R^2 proposed by Cohen (1988), Wetzels et al. (2009) derived cutoff values for small, medium, and large effect sizes of R^2 (GoF_{small} = 0.1; GoF_{medium} = 0.25; GoF_{large} = 0.36). In our study, the average communality of all major constructs was 0.684 and the average R^2 of the endogenous latent variables was 0.364, thus yielding a GoF value of 0.499, exceeding the cutoff value for large effect sizes of R^2 (Wetzels et al., 2009). Overall, the main model exhibited good theoretical explanatory power for the we-intention to participate in collective trolling.

To test the mediation hypotheses, we used the bootstrap test, as recommended by Preacher et al., (2007) and Zhao et al., (2010), to check the significance level of the indirect effects. As shown in Model 3 in Table 5, after bootstrapping 1,000 resamples, the indirect effect of the anonymity of self on the we-intention to participate in collective trolling via perceived online disinhibition was significant (β = 0.131), with bias-corrected 95% confidence intervals (BC 95% CI) between 0.071 and 0.242. The indirect effect of the anonymity of others on the we-intention to participate in collective trolling via social identity was also significant (β = 0.132; BC 95% CI, 0.066 to 0.193). The direct effects of the anonymity of self (β = 0.130; $p > 0.05$) and the anonymity of others (β = 0.097; $p > 0.05$) on the we-intention to participate in collective trolling were insignificant (see Model 3 in Table 5). These results show that perceived online disinhibition fully mediated the effect of the anonymity of self on the we-intention to participate in collective trolling, whereas social identity fully mediated the effect of the anonymity of others on the we-intention to participate in collective trolling. Therefore, H1 and H2 were supported.

To test the moderated mediation hypotheses, we further added the moderator (i.e., the absence of capable guardianship) and two associated interaction terms to the model. The PLS-SEM results show that the absence of capable guardianship strengthened the positive effect of perceived online disinhibition on the we-intention to participate in collective trolling (β = 0.289; $p < 0.05$) but had no significant effect on the relationship between social identity and the we-intention to participate in collective trolling (β = 0.032; $p > 0.05$) (see Model 4 in Table 5 and Figure 2).

Table 5. PLS-SEM Results of Mediation and Moderated Mediation Effects Test

	Model 1	Model 2	Model 3	Model 4
Path	***β***	***β***	***β***	***β***
Block 1: control variables				
Age → WPCT	−0.048	−0.009ns	0.034ns	0.045ns
Gender → WPCT	−0.012	−0.036ns	−0.078ns	−0.050ns
Education → WPCT	0.254	0.246***	0.192**	0.148*
Community experience → WPCT	−0.004	−0.030ns	−0.052ns	−0.004ns
Block 2: mediating mechanism				
AS → WPCT		0.315***	0.130ns	0.100ns
AO → WPCT		0.236***	0.097ns	0.062ns
AS → POD			0.580***	0.580***
AO → SI			0.472***	0.472***
POD → WPCT			0.225*	0.191*
SI → WPCT			0.280**	0.282**
Block 3: moderating mechanism				
ACG → WPCT				0.220**
ACG*POD → WPCT				0.289*
ACG*SI → WPCT				0.032ns
Indirect effect				
AS → POD → WPCT			0.131*	0.111*
AO → SI →WPCT			0.132*	0.133*
Index of moderated mediation				
AS → POD → WPCT on ACG				0.168*
AO → SI →WPCT on ACG				0.015ns
R^2	0.061	0.303	0.421	0.533
ΔR^2		0.242	0.118	0.112
f^2*-statistics*		0.347	0.204	0.240

Note: ACG = absence of capable guardianship, AO = anonymity of others, AS = anonymity of self, POD = perceived online disinhibition, SI = social identity, WPCT = we-intention to participate in collective trolling. *p < 0.05, **p < 0.01, ***p < 0.001, ns = not significant. f^2-statistics = $(R_{AB}^2 - R_A^2)/(1-R_{AB}^2)$, where R_A^2 is the variance explained by a set of independent variables A, and R_{AB}^2 is the combined variance explained by A and another set of independent variables B (Cohen, 1988).

According to Preacher et al. (2007) and Hayes (2015), our proposed research model is a second-stage moderated mediation model: the index of moderated mediation is the product value of a_1b_3, where a_1 is the estimated coefficient between the independent variable and the mediator and b_3 is the estimated coefficient of the interaction term. We consistently used the bootstrapping method with 1,000 iterations to test the significance level of the index of moderated mediation (Preacher et al., 2007; Hayes, 2015). The results in Table 5 (see Model 4) show that the index of moderated mediation exerted by the absence of capable guardianship on the mediating role of perceived online disinhibition was significant (β = 0.168; BC 95% CI: [0.065, 0.267]), thus supporting H3a. However, the index of moderated mediation exerted by the absence of capable guardianship on the mediating role of social identity was not significant (β = 0.015; BC 95% CI: [−0.056, 0.066]). Therefore, H3b was not supported. Moreover, the results in Table 5 (ΔR^2 = 0.112, f^2-statistic = 0.240 between Model 4 and Model 3) indicate that incorporating the moderating role of the absence of capable guardianship significantly improved the predictive power of the research model with a moderate to large effect size (Cohen, 1988).

To ensure the robustness of the findings, we further cross-validated the results using the PROCESS macro in SPSS. This approach integrates the causal steps approach (Baron & Kenny, 1986), the normal theory approach (i.e., the Sobel test) (Sobel, 1982), and the bootstrapping approach (Hayes, 2017). As such, the PROCESS macro approach provides a holistic and robust test of mediation and moderated mediation effects and has been widely used in the literature (Zhao et al., 2010; Chan et al., 2019; Wong et al., 2021). We first standardized all of the data to eliminate potential multicollinearity problems and avoid biased estimates of coefficients (Hayes, 2017). All of the latent constructs were measured by averaging their associated items. We included all of the control variables and the other constructs as covariates in the analysis of each mediation and moderated mediation effect.

Table 6. Regression Results of Mediation and Moderated Mediation Test in PROCESS Macro

	AS→POD→WPCT: BC 95% CI				AO→SI→WPCT: BC 95% CI			
	β	**SE**	**LLCI**	**ULCI**	***β***	**SE**	**LLCI**	**ULCI**
Mediation effect test:								
Indirect effect	0.122	0.050	0.022	0.224	0.122	0.043	0.045	0.215
Direct effect	0.123	0.052	0.021	0.224	0.091	0.049	−0.005	0.187
Moderated mediation effect test: the level of moderator (ACG)								
Low (−1 SD)	-0.008	0.067	-0.107	0.095	0.119	0.044	0.043	0.218
High (+1 SD)	0.212	0.069	0.107	0.333	0.142	0.067	0.025	0.282
Difference	0.220	0.094	0.060	0.346	0.023	0.071	-0.083	0.148
Index of moderated mediation	0.132	0.056	0.036	0.208	0.014	0.044	-0.051	0.091

Note: LLCI/ULCI = lower/upper limit of confidence interval, SE = standard error, AS = anonymity of self, AO = anonymity of others, ACG = absence of capable guardianship, POD = perceived online disinhibition, SI = social identity, WPCT = we-intention to participate in collective trolling. BC 95% CI = bias-corrected 95% confidence intervals. The number of bootstrap samples is 5000.

Table 7. Summary of Hypothesis Testing Results

Hypotheses	**Results**	**Supported?**
H1: Perceived online disinhibition mediates the effect of the anonymity of self on the we-intention to participate in collective trolling in virtual communities.	**PLS-SEM:** Indirect effect: 0.131* Direct effect: 0.130ns **PROCESS:** Indirect effect: 0.122* Direct effect: 0.123*	Yes
H2: Social identity mediates the effect of the anonymity of others on the we-intention to participate in collective trolling in virtual communities.	**PLS-SEM:** Indirect effect: 0.132* Direct effect: 0.097ns **PROCESS:** Indirect effect: 0.122* Direct effect: 0.091ns	Yes
H3a: The absence of capable guardianship moderates the mediating effect of perceived online disinhibition between the anonymity of self and the we-intention to participate in collective trolling such that the mediating effect will be stronger when the absence of capable guardianship is high compared to when it is low.	**Index of moderated mediation:** **PLS-SEM:** 0.168* **PROCESS:** 0.132*	Yes
H3b: The absence of capable guardianship moderates the mediating effect of social identity between the anonymity of others and the we-intention to participate in collective trolling such that the mediating effect will be stronger when the absence of capable guardianship is high compared to when it is low.	**Index of moderated mediation:** **PLS-SEM:** 0.015ns **PROCESS:** 0.014ns	No

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ns=not significant.

Table 6 shows that perceived online disinhibition partially mediated the effect of the anonymity of self on the we-intention to participate in collective trolling, whereas social identity fully mediated the effect of the anonymity of others on the we-intention to participate in collective trolling, thus confirming H1 and H2. The absence of capable guardianship positively moderated the indirect effect of the anonymity of self on the we-intention to participate in collective trolling via perceived online disinhibition, but it did not moderate the mediating effect of social identity on the relationship between the anonymity of others and the we-intention to participate in collective trolling. Therefore, H3a was supported, whereas H3b was not supported, consistent with the results of the PLS-SEM analysis. Table 7 summarizes the hypothesis testing results.

6 Discussion and Implications

This study seeks to theorize collective trolling in virtual communities. Specifically, we addressed our two research questions by introducing the concept of *we-intention* to capture the collective nature of collective trolling participation in virtual communities. We also used the sociotechnical perspective by integrating the SIDE model and the situational action theory to develop a model to explain the we-intention to participate in collective trolling in virtual communities. The research model was tested with online responses obtained from 377 Reddit users, and most of the hypotheses were supported.

6.1 Discussion of Key Findings

Our results provide empirical evidence for the two major mechanisms of the SIDE model: (1) perceived online disinhibition mediates the effect of the anonymity of self on the we-intention to participate in collective trolling; and (2) social identity mediates the effect of the anonymity of others on the we-intention to participate in collective trolling. Consistent with the sociotechnical perspective (Wixom & Todd, 2005; Sarker et al., 2019), the key role of individual-based social elements (i.e., perceived online disinhibition and social identity) channels and internalizes the technical elements of virtual communities (i.e., anonymity of self and anonymity of others) into behavioral intention (i.e., the we-intention to participate in collective trolling).

Situational action theory highlights the person-environment interaction effect on deviant behaviors. Our results show that the guardianship environment determines how the anonymity of self influences the we-intention to participate in collective trolling via perceived online disinhibition. When capable guardianship is established to deter collective trolling in a virtual community, community members who feel disinhibited due to the anonymity of self may be deterred from engaging in collective trolling in that community. This suggests that capable guardianship can counterbalance the unintended negative impact of the anonymity of self, which is a technical design choice made by virtual community owners and can foster the we-intention to participate in collective trolling by triggering perceived online disinhibition.

However, contrary to our expectations, the influence of the anonymity of others on the we-intention to participate in collective trolling via social identity is not subject to the guardianship environment. A possible reason for this finding is that groupthink occurs when community members feel that they are part of a group that they cherish belonging to (i.e., have a strong sense of social identity) (Turner & Pratkanis, 1998). This mode of groupthink leads individuals to believe unquestioningly in the inherent morality of their group and thus underestimate the ethical or moral consequences of their actions (e.g., collective trolling) (Janis, 1991; Krumsiek, 2017). As a result, the effect of guardianship appears to be weaker in the context of collective trolling. Moreover, studies have shown that the perceived risk of sanctions and perceived accountability for deviant behaviors are reduced when individuals believe that a group of people are involved in the deviant action (McGloin & Thomas, 2016; Alnuaimi et al., 2010). Accordingly, members who define themselves as part of the group are less likely to be deterred by guardianship in the context of collective trolling in virtual communities. Overall, the results of our moderated mediation analysis provide partial support for situational action theory (Wikström, 2004, 2014) by revealing that the person-environment interactive effect on deviant behaviors varies across individual-based social elements (i.e., perceived online disinhibition and social identity) under the same environmental conditions (i.e., the absence of capable guardianship).

6.2 Theoretical Implications

Our study is one of the first (if not the first) academic studies to theorize collective trolling in virtual communities. Understanding why members of virtual communities participate in coordinated trolling campaigns is critical to address concerns about the rise in online collective deviant behaviors afforded by social technologies (Ransbotham et al., 2016; Krumsiek, 2017; Al-khateeb & Agarwal, 2019). The current study complements practice by enriching the theoretical understanding of collective trolling in virtual communities. Accordingly, this study has a number of significant implications for IS research.

First, online deviant behaviors have been attracting increasing attention from the IS community (Ransbotham et al., 2016; Lowry et al., 2016; Venkatraman et al., 2018; Chan et al., 2019). Our study enriches IS research by focusing on a unique and serious form of online deviant behavior—collective trolling. Collective trolling is a widespread threat in virtual communities but has received minimal attention from IS researchers: the literature on this phenomenon comprises only a few descriptive analyses (Flores-Saviaga et al., 2018; Kirkwood et al., 2019; Sun & Fichman, 2020). Given the collective nature of this form of trolling, we introduce the concept of we-intention to conceptualize members' participation in collective trolling as a group-referent intentional action. Compared with constructs measuring individual participation in online deviant behaviors, the concept of we-intention offers a better theorization of participation in collective trolling in virtual communities because community members refer to other members when making decisions. We-intention has long been recognized as a useful concept to explain a group-referent intentional action in the IS literature (Bagozzi 2007; Tsai & Bagozzi, 2014; Shen et al., 2010, 2011, 2021). However, previous work tends to associate we-intention with positive behaviors, such as knowledge sharing, using social media for collective action, and playing online games, with little understanding of the concept in the context of online deviant behaviors (see Table A1 in Appendix). Our theoretical explanation and empirical investigation of the we-intention to participate in collective trolling are expected to stimulate and inform additional IS discourses on this emerging but underexplored phenomenon. We have also extended the group-referent intentional action literature to the online deviant behavior context.

Second, we adopted a sociotechnical perspective to explore collective trolling as a sociotechnical phenomenon and developed a research model that explains the we-intention to participate in collective trolling in virtual communities. The sociotechnical perspective comprises a major focus in the IS field (Sarker et al. 2019). By developing a model of the we-intention to participate in collective trolling, our work represents the first attempt to leverage the potential of IS research to theorize and develop ways to combat collective trolling in virtual communities. Specifically, we integrate the SIDE model and situational action theory to invoke a sociotechnical perspective for explaining the we-intention to participate in collective trolling in virtual communities. The SIDE model highlights two mechanisms of anonymity effects on collective actions in online groups, which is relevant to our theorization from the sociotechnical perspective. Following situational action theory, we further consider the moderating role of the absence of capable guardianship, thus incorporating the other important facet of social elements (i.e., environment-based social elements) into the sociotechnical framework. Our results demonstrate the considerable explanatory power of our proposed research model to explain the we-intention to participate in collective trolling in virtual communities. Therefore, our research model offers a good starting point for IS researchers to explain the we-intention to participate in collective trolling in virtual communities.

Third, our study contributes to research on the SIDE model and situational action theory. The original SIDE model is not concerned with anonymity and does not treat it as a defining feature (Spears, 2017). In this study, we advanced the SIDE model by identifying two distinctive roles of anonymity (anonymity of self and anonymity of others) that capture the online representation of individuals and groups in virtual communities. Further, our findings empirically confirm two major assumptions of the SIDE model by demonstrating the mediating role of perceived online disinhibition between the anonymity of self and the we-intention to participate in collective trolling as well as the mediating role of social identity between the anonymity of others and the we-intention to participate in collective trolling. The application of situational action theory to examine the moderating influence of the absence of capable guardianship on the two mechanisms of SIDE also specifies the theoretical boundaries within which the SIDE model works, thereby advancing research on the SIDE model. Moreover, our findings suggest that the absence of capable guardianship acts as a positive moderator of the strategic mechanism of SIDE but fails to moderate the cognitive mechanism of SIDE. These two different moderating roles of the absence of capable guardianship pose a challenge to the assumption of situational action theory in that the we-intention to participate in collective trolling is not necessarily the result of the interaction between individual-based and environment-based social elements. These findings pave the way for future studies to delve into the applicability and contextualization of situational action theory across contexts with various individual-based and environment-based social elements.

6.3 Practical Implications

The findings of this study have important implications for managerial practice. Our sociotechnical model shows that the two types of anonymity offered by virtual communities may both play fundamental roles in the formation of the we-intention to participate in collective trolling in virtual communities via perceived online disinhibition and social identity. Therefore, virtual community owners could manipulate anonymity features to change members' psychological and motivational states (e.g., perceived online disinhibition and social identity) and thereby intervene in collective trolling in virtual communities.

Specifically, to change members' perceptions of the anonymity of self, virtual community platforms could ask members to provide sensitive personal information (e.g., name, email address, and phone number) upon registration on the premise of ensuring security and confidentiality. Gathering such information would make members of virtual communities behave in more accountable and inhibited ways, reducing the we-intention to participate in collective trolling. Platforms could also encourage users to create personalized usernames, including their job title, gender, location, or other personalized information upon registration and prevent users from frequently changing their IDs or usernames, thereby making their personal profiles consistent and long-lasting. As a result, members would regard themselves as more easily identifiable during online interactions, making them feel inhibited and less likely to form a we-intention to participate in collective trolling in virtual communities. Long-lasting individuating information on members would also allow community members to distinguish other members, thereby reducing the anonymity of others (Ma & Agarwal, 2007). Therefore, the we-intention to participate in collective trolling would also be reduced.

Virtual community platforms could also reduce the anonymity of others by allowing their members to check other members' chatting/posting history (i.e., deep profiling) (Ma & Agarwal, 2007). Virtual community platforms could provide navigation or search tools to allow members to search for posts by particular members, obtain individuating information on focal members (e.g., their expertise and social networks), and categorize members in unique ways. This strategy would weaken community members' social identity salience, making them less likely to form we-intentions to participate in collective trolling in virtual communities.

Furthermore, we find that the absence of capable guardianship acts as a positive moderator of the mediating role of perceived online disinhibition between the anonymity of self and the we-intention to participate in collective trolling. When capable guardianship is established to deter collective trolling, disinhibited members are less likely to use the strategic advantages provided by the anonymity of self to engage in collective trolling to avoid punishment. However, the absence of capable guardianship does not affect the mediating role of social identity between the anonymity of others and the we-intention to participate in collective trolling because community members with a strong sense of social identity are likely to ignore or underestimate the deterrence effect of guardianship. Accordingly, to increase the effectiveness of guardianship in deterring potential trollers, platforms could issue zero-tolerance policies with clear punishments for collective trollers. Collective trolling should also be clearly defined and scoped, and any violation of policies and rules prohibiting collective trolling should be punished, regardless of the number of people involved. Platforms could establish collective trolling detection and warning systems. They could also provide incentives and design reputation systems to motivate community moderators to safeguard communities against collective trolling. Training programs and individual mentorship would also help empower community moderators to resist collective trolling in virtual communities.

6.4 Limitations and Directions for Future Research

Although our study has important implications for research and practice, its contributions should be interpreted in light of the following limitations, which suggest several opportunities for future research. First, we only considered the effects of two types of anonymity on the we-intention to participate in collective trolling in virtual communities. Studies have shown increasing interest in understanding the ways that social technology facilitates the organization of online collective action (e.g., Vaast et al., 2017; Sæbø et al., 2020). Accordingly, future research could develop a holistic sociotechnical model by considering the effects of additional social technology features on online collective actions (especially collective deviant actions). In particular, future studies could further explore how the interplay between social components and technical components jointly influences online collective actions by identifying social or technical boundary conditions. Another possible research avenue related to this opportunity would be to investigate how the selection and configuration of these social technology features may contribute to the organization of collective actions.

Second, as the core construct in our model, anonymity was often operationalized as a unidimensional construct in prior studies. However, anonymity is a complex concept that takes various forms, such as content vs. process anonymity, social vs. technical anonymity, and self-to-others vs. others-to-self anonymity (McLeod, 1997; Pinsonneault & Heppel, 1997; Lowry et al., 2016). Because of our research focus, we built on the SIDE model and conceptualized anonymity in terms of the anonymity of self and the anonymity of others to capture the online representation of individuals and groups in virtual communities. We strongly recommend that future studies conceptualize and operationalize anonymity in different ways to investigate the differential impact of anonymity on collective actions.

Third, we drew upon the SIDE model to examine the two distinctive roles of anonymity in the context of virtual community: the anonymity of self via perceived online disinhibition (i.e., the strategic mechanism of the SIDE model) and the anonymity of others via social identity (i.e., the cognitive mechanism of the SIDE model). While prior studies have identified and examined various mechanisms (e.g., group conformity, depersonalization, social support, and self-awareness) in the SIDE model (Lea et al., 2001; Lee, 2004; Spears et al., 2002; Kim et al., 2019), future research could continue to explore the mechanisms between anonymity and collective trolling. Furthermore, the strategic mechanism captures the effects of identifiability and accountability based on the assumption that individuals are sensitive to the power or socially accepted norms in a virtual community. Although we anticipate expect that our respondents were aware of in-group and out-group members on Reddit, future studies could manipulate the identity level of audience groups (e.g., powerful out-groups, authorities, and in-groups) and examine its impact on members' decision to participate in collective trolling.

Fourth, we validated the proposed model using the online responses of users of a single social platform (i.e., Reddit). Although Reddit offered an appropriate setting for our study of anonymity and collective trolling, the generalizability of our sample to other social technology users should be carefully considered. However, our participants had heterogeneous demographic characteristics, such as education, age, gender, employment status, and income, which should help to mitigate concerns about the sampling procedure. Future studies could replicate our research model in other contexts to test the generalizability of our findings.

Finally, we used a cross-sectional self-reported survey to collect data to validate our model. As a result, our findings may have been influenced by response bias. We used several procedural and statistical measures to

detect and mitigate response bias. For example, we used a third-party platform to collect our data and also emphasized that the survey was entirely anonymous to encourage the participants to respond honestly. We also tested for common method bias and social desirability bias and found that neither contaminated our findings. Nevertheless, the survey approach generally depends on participants' voluntary responses, making it difficult to address self-selection bias. Future studies should use mixed methods designs, such as a combination of interviews, case studies, archival data analysis, and online surveys, to cross-validate our research findings. However, researchers using alternative research designs (e.g., interviews) should also ensure confidentiality and seek to reduce social desirability bias, which is likely to be a challenge for interviews in particular.

7 Conclusion

With the rise of virtual communities, collective trolling continues to attract increasing attention from academics and practitioners. This study introduces the concept of *we-intention* to theorize participation in collective trolling as a group-referent intentional action and further considers collective trolling from the sociotechnical perspective. Specifically, we integrate the SIDE model and situational action theory to invoke a sociotechnical perspective explaining the we-intention to participate in collective trolling in virtual communities. The results shed light on the mechanisms and boundary conditions of the influence of two anonymity features of virtual communities on the we-intention to participate in collective trolling. This study advances the theoretical understanding of online collective deviant behaviors facilitated by social technologies and offers practical guidance for virtual community owners on the formulation of timely measures to combat collective trolling.

Acknowledgments

We gratefully appreciate the comments and suggestions of the senior editor, Likoebe Maruping, and the anonymous reviewers of this paper throughout this paper's review process. Their excellent input has significantly improved the quality of our work. This work was substantially supported by a fellowship award from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKBU SRFS2021-2H03), and partially supported by the grants from the National Natural Science Foundation of China (Project No. 71828202, 71832010).

References

Al-Khateeb, S., & Agarwal, N. (2014). Developing a conceptual framework for modeling deviant cyber flash mob: A socio-computational approach leveraging hypergraph constructs. *Journal of Digital Forensics, Security and Law*, *9*(2), Article 10.

Al-khateeb, S., & Agarwal, N. (2019). *Deviance in social media and social cyber forensics: Uncovering hidden relations using open source information (OSINF)*. Springer.

Alnuaimi, O. A., Robert, L. P., & Maruping, L. M. (2010). Team size, dispersion, and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems*, *27*(1), 203-230.

Bagozzi, R. P. (2000). On the concept of intentional social action in consumer behavior. *Journal of Consumer Research*, *27*(3), 388-396.

Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, *8*(4), 244-254.

Bagozzi, R. P., & Dholakia, U. M. (2006a). Open source software user communities: A study of participation in Linux user groups. *Management Science*, *52*(7), 1099-1115.

Bagozzi, R. P., & Dholakia, U. M. (2006b). Antecedents and purchase consequences of customer participation in small group brand communities. *International Journal of Research in Marketing*, *23*(1), 45-61.

Bagozzi, R. P., & Dholakia, U. M. (2002). Intentional social action in virtual communities. *Journal of Interactive Marketing*, *16*(2), 2-21.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Berghel, H., & Berleant, D. (2018). The online trolling ecosystem. *IEEE Computer*, *51*(8), 44-51.

Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and failures: A sociotechnical perspective part I: The cause. *MIS Quarterly*, *1*(3), 17-32.

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review*, *101*(2), 327-341.

Chan, T. K. H., Cheung, C. M. K., & Lee, Z. W. Y. (2021). Cyberbullying on social networking sites: A literature review and future research directions. *Information & Management*, Article 103411.

Chan, T. K. H., Cheung, C. M. K., & Wong, R. Y. M. (2019). Cyberbullying on social networking sites: The crime opportunity and affordance perspectives. *Journal of Management Information Systems*, *36*(2), 574-609.

Chen, J. V., Hiele, T. M., Kryszak, A., & Ross, W. H. (2020). Predicting intention to participate in socially responsible collective action in social networking website groups. *Journal of the Association for Information Systems*, *21*(2), 341-363.

Cheung, C. M. K., Chiu, P. Y., & Lee, M. K. O. (2011). Online social networks: Why do students use Facebook? *Computers in Human Behavior*, *27*(4), 1337-1343.

Cheung, C. M. K., & Lee, M. K. O. (2010). A theoretical model of intentional social action in online social networks. *Decision Support Systems*, *49*(1), 24-30.

Cheung, C. M. K, Wong, R. Y. M., & Chan, T. K. H. (2021). Online disinhibition: conceptualization, measurement, and implications for online deviant behavior. *Industrial Management & Data Systems*.

Citron, D. K. (2014). *How cyber mobs and trolls have ruined the internet—and destroyed lives.* Newsweek. https://www.newsweek.com/internet-and-golden-age-bully-271800

Citron, D. K. (2020). Cyber Mobs, Disinformation, and death videos: The Internet as it is (and as it should be). *Michigan Law Review*, *118*(6), 1073-1093.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum.

Coles, B. A., & West, M. (2016). Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior*, *60*, 233-244.

Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, *102*, 79-84.

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

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A

deterrence approach. *Information Systems Research*, *20*(1), 79-98.

de Oliveira, M. J., & Huertas, M. K. Z. (2015). Does life satisfaction influence the intention (We-Intention) to use Facebook? *Computers in Human Behavior*, *50*, 205-210.

Dholakia, U. M., Bagozzi, R. P., & Pearo, L. K. (2004). A social influence model of consumer participation in network-and small-group-based virtual communities. *International Journal of Research in Marketing*, *21*(3), 241-263.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335-358.

Dineva, D., & Breitsohl, J. (2021). Managing trolling in online communities: An organizational perspective. *Internet Research*, *32*(1), 292-311.

Durcikova, A., Lee, A. S., & Brown, S. A. (2018). Making rigorous research relevant: Innovating statistical action research. *MIS Quarterly*, *42*(1), 241-263.

Felson, M., & Clarke, R. V. (1998). *Opportunity makes the thief: Practical theory for crime prevention* (Police Research Group Paper 98). London: Home Office; Research, Development and Statistics Directorate.

Ferguson, C. J., & Glasgow, B. (2021). Who are GamerGate? A descriptive study of individuals involved in the GamerGate controversy. *Psychology of Popular Media*, *40*(2), 243-247.

Flores-Saviaga, C. I., Keegan, B. C., & Savage, S. (2018). Mobilizing the Trump train: Understanding collective action in a political trolling community. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media* (pp. 82-91).

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50.

Gabbiadini, A., Mari, S., & Volpato, C. (2013). Virtual users support forum: Do community members really want to help you? *Cyberpsychology, Behavior, and Social Networking*, *16*(4), 285-292.

Gilbert, M. (1989). *On social facts.* Routledge.

Gutman, R. (2018). *Reddit's case for anonymity on the Internet.* The Atlantic. https://www.theatlantic.com/technology/archive/2018/06/reddit-anonymity-privacy-authenticity/564071/

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139-152.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, *6*(2), 215-242.

Harrison, A. (2018). The effects of media capabilities on the rationalization of online consumer fraud. *Journal of the Association for Information Systems*, *19*(5), 408-440.

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1-22.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford.

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, *28*(2), 565-580.

Huang, K. Y., Chengalur-Smith, I., & Pinsonneault, A. (2019). Sharing is caring: Social support provision and companionship activities in healthcare virtual support communities. *MIS Quarterly*, *43*(2), 395-424.

Huang, N., Hong, Y., & Burtch, G. (2017). Social network integration and user content generation: Evidence from natural experiments. *MIS Quarterly*, *41*(4), 1035-1058.

Janis, I. L. (1991). Groupthink. In E. Griffin (Ed.) *A First Look at Communication Theory* (pp. 235-246). McGrawHill.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199-218.

Jiang, Z., Heng, C. S., & Choi, B. C. (2013). Research note—privacy concerns and privacy-protective behavior in synchronous online social interactions. *Information Systems Research*, *24*(3), 579-595.

Karahanna, E., Xu, S. X., Xu, Y., & Zhang, N. A. (2018). The needs-affordances-features perspective for the use of social media. *MIS Quarterly*, *42*(3), 737-756.

Kling, R., & Courtright, C. (2003). Group behavior and learning in electronic forums: A sociotechnical approach. *The Information Society*, *19*(3), 221-235.

Kim, K. K., Lee, A. R., & Lee, U. K. (2019). Impact of anonymity on roles of personal and group identities in online communities. *Information & Management*, *56*(1), 109-121.

Kirkwood, G. L., Payne, H. J., & Mazer, J. P. (2019). Collective trolling as a form of organizational resistance: Analysis of the# Justiceforbradswife twitter campaign. *Communication Studies*, *70*(3), 332-351.

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

Kordzadeh, N., & Warren, J. (2017). Communicating personal health information in virtual health communities: An integration of privacy calculus model and affective commitment. *Journal of the Association for Information Systems*, *18*(1), 45-81.

Krumsiek, A. (2017). *Cyber mobs: Destructive online communities*. Greenhaven.

Kwak, D. H., Holtkamp, P., & Kim, S. S. (2019). Measuring and controlling social desirability bias: Applications in information systems research. *Journal of the Association for Information Systems*, *20*(4), 317-345.

Laato, S., Inaba, N., Paloheimo, M., & Laajala, T. D. (2021). Group polarisation among location-based game players: An analysis of use and attitudes towards game slang. *Internet Research*, *31*(5), 1695-1717.

Lea, M., Spears, R., & de Groot, D. (2001). Knowing me, knowing you: Anonymity effects on social identity processes within groups. *Personality and Social Psychology Bulletin*, *27*(5), 526-537.

Lee, E. J. (2004). Effects of visual representation on social influence in computer-mediated communication: Experimental tests of the social identity model of deindividuation effects. *Human Communication Research*, *30*(2), 234-259.

Lew, L. (2019). *Hong Kong protests and 'fake news': In the psychological war for hearts and minds, disinformation becomes a weapon used by both sides.* South China Morning Post. https://www.scmp.com/news/hong-kong/society/article/3032734/fake-news-and-hong-kong-protests-psychological-war-hearts

Li, M., & Suh, A. (2021). We-intention to continue playing mobile multiplayer games: The role of social play habit. *Internet Research*, *31*(4), pp. 1153-1176.

Liang, H., Xue, Y., Pinsonneault, A., & Wu, Y. A. (2019). What users do besides problem-focused coping when facing IT security threats: An emotion-focused coping perspective. *MIS Quarterly*, *43*(2), 373-394.

Linvill, D. L., Boatwright, B. C., Grant, W. J., & Warren, P. L. (2019). "THE RUSSIANS ARE HACKING MY BRAIN!" Investigating Russia's internet research agency Twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior*, *99*, 292-300.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*(1), 114-114.

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, *27*(4), 962-986.

Lowry, P. B., Zhang, J., Moody, G. D., Chatterjee, S., Wang, C., & Wu, T. (2019). An integrative theory addressing cyberharassment in the light of technology-based opportunism. *Journal of Management Information Systems*, *36*(4), 1142-1178.

Ma, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, *18*(1), 42-67.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293-334.

Majchrzak, A., & Malhotra, A. (2016). Effect of knowledge-sharing trajectories on innovative outcomes in temporary online crowds. *Information Systems Research*, *27*(4), 685-703.

Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for digital transformation: Lessons for information systems research from the study of ICT and societal challenges. *MIS Quarterly*, *40*(2), 267-277.

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science*, *52*(12), 1865-1883.

Maruping, L. M., Daniel, S. L., & Cataldo, M. (2019). Developer centrality and the impact of value congruence and incongruence on commitment and code contribution activity in open source software communities. *MIS Quarterly*, *43*(3), 951-976.

Massanari, A. (2017). # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, *19*(3), 329-346.

Masui, K. (2019). Loneliness moderates the relationship between Dark Tetrad personality traits and internet trolling. *Personality and Individual Differences*, *150*, 109475.

McGloin, J. M., & Thomas, K. J. (2016). Incentives for collective deviance: Group size and changes in perceived risk, cost, and reward. *Criminology*, *54*(3), 459-486.

McLeod, P. (1997). A comprehensive model of anonymity in computer-supported group decision making. *Proceedings of the Eighteenth International Conference on Information Systems* (pp. 223-234).

Mindel, V., Mathiassen, L., & Rai, A. (2018). The sustainability of polycentric information commons. *MIS Quarterly*, *42*(2), 607-632.

Morschheuser, B., Riar, M., Hamari, J., & Maedche, A. (2017). How games induce cooperation? A study on the relationship between game features and we-intentions in an augmented reality game. *Computers in Human Behavior*, *77*, 169-183.

Ortiz, S. M. (2020). Trolling as a collective form of harassment: An inductive study of how online users understand trolling. *Social Media+ Society*, *6*(2), Article 2056305120928512.

Pauwels, L. J., Svensson, R., & Hirtenlehner, H. (2018). Testing situational action theory: A narrative review of studies published between 2006 and 2015. *European Journal of Criminology*, *15*(1), 32-55.

Pinsonneault, A., & Heppel, N. (1997). Anonymity in group support systems research: A new conceptualization, measure, and contingency framework. *Journal of Management Information Systems*, *14*(3), 89-108.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, *123*(3), 238-259.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185-227.

Ransbotham, S., Fichman, R. G., Gopal, R., & Gupta, A. (2016). Special section introduction—ubiquitous IT and digital vulnerabilities. *Information Systems Research*, *27*(4), 834-847.

Ransbotham, S., & Mitra, S. (2009). Choice and chance: A conceptual model of paths to information security compromise. *Information Systems Research*, *20*(1), 121-139.

Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European review of Social Psychology*, *6*(1), 161-198.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quarterly*, *36*(3), 841-864.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*(1), 119-125.

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Addison-Wesley.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, *12*(4), 762-800.

Sæbø, Ø., Federici, T., & Braccini, A. M. (2020). Combining social media affordances for organising collective action. *Information Systems Journal*, *30*(4), 699-732.

Sanfilippo, M. R., Fichman, P., & Yang, S. (2018). Multidimensionality of online trolling behaviors. *The Information Society*, *34*(1), 27-39.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695-720.

Sarstedt, M., Hair Jr, J. F., Nitzl, C., Ringle, C. M., & Howard, M. C. (2020). Beyond a tandem analysis of SEM and PROCESS: Use of PLS-SEM for mediation analyses! *International Journal of Market Research*, *62*(3), 288-299.

Sattelberg, W. (2020). *The demographics of Reddit: Who uses the site?* TechJunkie. https://social.techjunkie.com/demographics-reddit/

Scott, S. V., & Orlikowski, W. J. (2014). Entanglements in practice: Performing anonymity through social media. *MIS Quarterly*, *38*(3), 873-894.

Shen, X. L., Cheung, C. M. K., Lee, M. K. O., & Chen, H. (2011). How social influence affects we-intention to use instant messaging: The moderating effect of usage experience. *Information Systems Frontiers*, *13*(2), 157-169.

Shen, X. L., Cheung, C. M. K., & Lee, M. K. O. (2013). Perceived critical mass and collective intention in social media-supported small group communication. *International Journal of Information Management*, *33*(5), 707-715.

Shen, X. L., Lee, M. K. O., & Cheung, C. M. K. (2012). Harnessing collective intelligence of Web 2.0: Group adoption and use of Internet-based collaboration technologies. *Knowledge Management Research & Practice*, *10*(4), 301-311.

Shen, X. L., Lee, M. K. O., & Cheung, C. M. K. (2014). Exploring online social behavior in crowdsourcing communities: A relationship management perspective. *Computers in Human Behavior*, *40*, 144-151.

Shen, X. L., Lee, M. K. O., Cheung, C. M. K., & Chen, H. (2010). Gender differences in intentional social action: We-intention to engage in social network-facilitated team collaboration. *Journal of Information Technology*, *25*(2), 152-169.

Shen, X. L., Li, Y. J., Sun, Y., Chen, J., & Wang, F. (2019). Knowledge withholding in online knowledge spaces: Social deviance behavior and secondary control perspective. *Journal of the Association for Information Science and Technology*, *70*(4), 385-401.

Shen, X. L., Li, Y. J., Sun, Y., & Wang, F. (2021). Good for use, but better for choice: A relative model of competing social networking services. *Information & Management*, *58*(3), Article 103448.

Sia, C. L., Tan, B. C., & Wei, K. K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research*, *13*(1), 70-90.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290-312.

Song, H., & Lee, S. S. (2020). Motivations, propensities, and their interplays on online bullying perpetration: a partial test of situational action theory. *Crime & Delinquency*, *66*(12), 1787-1808.

Spears, R. (2017). Social identity model of deindividuation effects. In P. Rössler, C. A. Hoffner, & L. van Zoonen (Ed.), *The international encyclopedia of media effects* (pp. 1-9). Wiley-Blackwell.

Spears, R., & Lea, M. (1994). Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research*, *21*(4), 427-459.

Spears, R., Lea, M., Corneliussen, R. A., Postmes, T., & Haar, W. T. (2002). Computer-mediated communication as a channel for social resistance: The strategic side of SIDE. *Small Group Research*, *33*(5), 555-574.

Spears, R., & Postmes, T. (2015). Group identity, social influence, and collective action online. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 23-46). Blackwell.

Springer, N. J. (2015). *Publics and counterpublics on the front page of the internet: The cultural practices, technological affordances, hybrid economics and politics of Reddit's public sphere* [Unpublished PhD dissertation]. University of Colorado at Boulder.

Srivastava, S. C., & Chandra, S. (2018). Social presence in virtual world collaboration: An uncertainty reduction perspective using a mixed methods approach. *MIS Quarterly*, *42*(3), 779-804.

Steelman, Z. R., Hammer, B. I., & Limayem, M. (2014). Data collection in the digital age: Innovative alternatives to student samples. *MIS Quarterly*, *38*(2), 355-378.

Strandberg, K., Himmelroos, S., & Grönlund, K. (2019). Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization. *International Political Science Review*, *40*(1), 41-57.

Suh, K. S., Lee, S., Suh, E. K., Lee, H., & Lee, J. (2018). Online comment moderation policies

for deliberative discussion-seed comments and identifiability. *Journal of the Association for Information Systems*, *19*(3), 182-208.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321-326.

Sun, H., & Fichman, P. (2018). Chinese collective trolling. *Proceedings of the Association for Information Science and Technology*, *55*(1), 478-485.

Sun, L. H., & Fichman, P. (2020). The collective trolling lifecycle. *Journal of the Association for Information Science & Technology*, *71*(7), 770-783.

Tajfel, H. (1978). Interindividual behavior and intergroup behavior. In H. Tajfel (Ed.), *Differentiation between social groups: Studies in the social psychology of intergroup relations* (pp. 27-60). Academic Press.

Turner, M. E., & Pratkanis, A. R. (1998). A social identity maintenance model of groupthink. *Organizational Behavior and Human Decision Processes*, *73*(2-3), 210-235.

Tsai, H. T., & Bagozzi, R. P. (2014). Contribution behavior in virtual communities: Cognitive, emotional, and social influences. *MIS Quarterly*, *38*(1), 143-164.

Tuomela, R. (1995). *The importance of us: A philosophical study of basic social notions*. Stanford, CA: Stanford University Press.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social media affordances for connective action: An examination of microblogging use during the Gulf of Mexico oil spill. *MIS Quarterly*, *41*(4), 1179-1205.

Venkatraman, S., Cheung, C. M. K., Lee, Z. W. Y., Davis, F. D., & Venkatesh, V. (2018). The "Darth" side of technology use: An inductively derived typology of cyberdeviance. *Journal of Management Information Systems*, *35*(4), 1060-1091.

Wang, C., & Lee, M. K. O. (2020). Why we cannot resist our smartphones: Investigating compulsive use of mobile SNS from a stimulus-response-reinforcement perspective. *Journal of the Association for Information Systems*, *21*(1), 175-200.

Wang, J., Shan, Z., Gupta, M., & Rao, H. R. (2019). A longitudinal study of unauthorized access attempts on information systems: The role of opportunity contexts. *MIS Quarterly*, *43*(2), 601-622.

Wang, N., & Sun, Y. (2016). Social influence or personal preference? Examining the determinants of usage intention across social media with different sociability. *Information Development*, *32*(5), 1442-1456.

Wetzels, M., Odekerken-Schröder, G., & van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, *33*(1), 177-195.

Wikström, P. O. H. (2004). Crime as alternative: Towards a cross-level situational action theory of crime causation. In J. McCord (Ed.), *Beyond empiricism: Institutions and intentions in the study of crime* (pp. 1-37). Transaction.

Wikström, P. O. H. (2014). Why crime happens: A situational action theory. In G. Manzo (Ed.), *Analytical sociology: Actions and networks* (pp. 74-94). Wiley.

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16*(1), 85-102.

Wong, R. Y. M., Cheung, C. M. K., & Xiao, B. (2018). Does gender matter in cyberbullying perpetration? An empirical investigation. *Computers in Human Behavior*, *79*, 247-257.

Wong, R. Y. M., Cheung, C. M. K., Xiao, B., & Thatcher, J. B. (2021). Standing up or standing by: Understanding bystanders' proactive reporting responses to social media harassment. *Information Systems Research*, *32*(2), 561-581.

Yang, Q., & Gong, X. (2021). The engagement-addiction dilemma: An empirical evaluation of mobile user interface and mobile game affordance. *Internet Research*, *31*(5), pp. 1745-1768

Yesiloglu, S., Memery, J., & Chapleo, C. (2021). To post or not to post? Examining motivations of brand related engagement types on social networking sites. *Internet Research*, *31*(5), 1849-1873.

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. *Companion Proceedings of the 2019 World Wide Web Conference* (pp. 218-226).

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197-206.

Appendix: Literature Review on We-intention

Methodology for Conducting Literature Review on We-Intention

Our literature search followed the standard guidelines of literature review proposed by Webster and Watson (2002). First, we searched a list of keywords "we-intention*" or "collective intention*" in the fields Title, Abstract, Author Keywords, and KeyWords Plus in the Web of Science. All electronic databases in Web of Science were included for the literature search. To ensure the comprehensiveness and relevance, we also searched the list of keywords "we-intention*" or "collective intention*" in the AIS Electronic Library that is the largest database in the IS field. We then applied inclusion criteria and exclusion criteria to these articles: (1) papers are journal articles written in English, (2) papers are theory-driven empirical research, (3) computer-mediated contexts, (4) "we-intention" is one of the research focuses. Finally, we obtained a total of 16 articles (as of May 2020) and Table A1 summarizes these articles.

Table A1. Empirical Research on "We-Intention" in the Literature

Authors (years)	Definition	Level of analysis	Theories	Key constructs	IT	DVs & contexts
Dholakia et al. (2004)	A commitment of an individual to engage in joint action and involves an implicit or explicit agreement between the participants to engage in that joint action (Tuomela, 1995).	Individual	U&G, SIT	Purposive value, self-discovery, maintaining interpersonal interconnectivity, social enhancement, entertainment value, group norms, mutual agreement, mutual accommodation, social identity	No	We-intentions to interact in small-group-based virtual communities
Bagozzi & Dholakia (2006a)	A special kind of intention in which the agent we-intends to perform an action jointly with the others or to see to it jointly with the others that a certain state comes about.	Individual	TPB, MGB	Attitude, perceived behavioral control, identification, positive or negative anticipated emotions, social identity	No	We-intentions to interact in Linux user groups in virtual communities
Bagozzi & Dholakia (2006b)	Tuomela's definition	Individual	TPB, MGB	Attitude, subjective norms, perceived behavioral control, positive or negative anticipated emotions, desire	No	We-intention to interact in small group brand communities
Cheung & Lee (2010)	Tuomela's definition	Individual	SIT	Subjective norm, group norm, social identity	No	We-intention to interact on Facebook
Cheung et al. (2011)	Tuomela's definition	Individual	U&G, SIT	Purposive value, self-discovery, maintaining interpersonal interconnectivity, social enhancement, entertainment value, subjective norm, group norm, social identity, social presence	No	We-intention to interact on Facebook
Shen et al. (2010)	An individual's subjective perception of the extent to which all participants in a collectivity will engage in a group activity together.	Individual	TRA, MGB, SIT	Attitude, positive or negative anticipated emotions, subjective norm, group norm, social identity	No	We-intention to use QQ groups to communicate
Shen et al. (2011)	Tuomela's definition	Individual	SIT, MGB	Subjective norm, group norm, social identity, desire	No	We-intention to use QQ groups

Shen et al. (2012)	One's perception of the group acting as a coordinated unit where members in the group collectively accept the action and commit themselves to performing this behavior.	Individual	U&G, TRA, MGB	Social entertainment, task accomplishment, social attention, meet new people, attitude, positive or negative anticipated emotions. Social identity	No	We-intention to use QQ groups
Shen et al. (2013)	An individual's subjective perception of the extent to which a particular group or social category will engage in a target collective behavior	Individual	CMT, SIT:	Perceived critical mass, subjective norm, group norm, social identity	No	We-intention to use QQ groups
Gabbiadini et al. (2013)	Tuomela's definition	Individual	TPB, SIT	Attitude, perceived behavioral control, group norms, social identity, free-riding tendency	No	We-intentions to contribute to virtual forums
Tsai & Bagozzi (2014)	Tuomela's definition	Individual	SIT, TPB, MGB	Subjective norm, group norm, social identity, attitude, perceived behavioral control, anticipated emotions, desire.]	No	We-intentions to contribute to virtual communities
Shen et al. (2014)	We-mode collective intention refers to acting as a group member, I-mode collective intention refers to acting interdependently to contribute to the group goal.	Individual	CTT	Team trust, commitment	No	We-intentions to contribute to Wiki communities
de Oliveira & Huertas (2015)	A special kind of intention where those involved intend to perform an action together with others	Individual	U&G, SIT	Purposive value, self-discovery, maintaining interpersonal interconnectivity, social enhancement, entertainment value, subjective norm, group norm, social identity, social presence, life satisfaction	No	We-intention to participate on Facebook
Wang & Sun (2016)	Tuomela's definition	Individual	SIT	Subjective norm, group norm, social identity, attitude	No	We-intention to interact on Weibo and Wechat
Morschheuser et al. (2017)	A "we-perspective" that expresses a collective commitment to participate in a cooperative action.	Individual	None	Engagement with cooperative or individualistic game features, group norms, positive or negative anticipated emotions, social identity, joint commitment, attitude	Yes	We-intention to play social games
Chen et al. (2020)	Tuomela's definition	Individual	SIT	Subjective norm, group norm, social identity, perceived corporate social responsibility	No	We-intention to use SNS for collective action

Note: U&G = use & gratification, SIT = social influence theory, TPB = theory of planned behavior, MGB = model of goal-directed behavior, TRA = theory of reasoned action, CMT = critical mass theory, CTT = commitment-trust theory.

About the Authors

Yang-Jun Li is a PhD candidate in the Department of Information Systems of City University of Hong Kong. His research interests include ethical issues in the use and application of IT, e-commerce, knowledge management, and social media. He has published in journals such as *Decision Support Systems*, *Journal of the Association for Information Science and Technology*, *Information & Management*, and *Journal of Business Research*.

Christy M. K. Cheung is a professor in the Department of Finance and Decision Sciences of Hong Kong Baptist University. She is an awardee of the Senior Research Fellow Scheme (Research Grants Council of the Hong Kong Special Administrative Region) with funding to advance research into the role of technology in online deviant behaviors. Her work appears in *Information Systems Research*, *Journal of the Association for Information Systems*, *Journal of Management Information Systems*, and *MIS Quarterly*. She serves as the editor-in-chief at *Internet Research*.

Xiao-Liang Shen is a professor in the School of Information Management of Wuhan University. His current research interests include information management, the dark side of IT, and digital governance. He has published in journals such as *Journal of the Association for Information Systems*, *Journal of Information Technology*, *Journal of the Association for Information Science and Technology*, *Information & Management*, and *Decision Support Systems*. He is also the corresponding author of this paper.

Matthew K. O. Lee is a chair professor of information systems and e-commerce in the Department of Information Systems at City University of Hong Kong. He is also the vice-president at the university. His research interests include IT-based innovation adoption and diffusion, knowledge management, electronic commerce, online addiction, and the development of digital competence. Professor Lee has published over one hundred refereed articles in international journals, such as *MIS Quarterly*, *Journal of MIS*, *Journal of AIS*, *International Journal of Electronic Commerce*, *Decision Support Systems*, *Information & Management*, and *Journal of International Business Studies*, among others. He has been rated as the most highly cited business professor at CityU for many years. Professor Lee has taken an editorial role in a number of leading scholarly journals including *MIS Quarterly*, *Information Systems Journal*, and *International Journal of Information Management*, among others.